

BAYESIAN PREDICTION IN MIXED LINEAR MODELS
WITH APPLICATIONS IN SMALL AREA ESTIMATION

BY

GAURI SANKAR DATTA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

UNIVERSITY OF FLORIDA LIBRARIES

1990

to my parents and teachers, with regards

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Malay Ghosh for being my advisor, for originally proposing the problem and for all the attention I received from him for the past five years. Without his enormous patience, encouragement and guidance, it would not have been possible to complete the work. Throughout my years in the graduate program, he has been my friend, philosopher and guide, and I consider myself extremely lucky to get him as my dissertation advisor. I would like to thank Professors Michael A. DeLorenzo and Ronald H. Randles for serving on my committee. Also, I am grateful to Professors Ramon C. Littell, Kenneth M. Portier and P.V. Rao for being on my Part C and oral defense committees.

My special thanks go to Professor Ghosh and Professor Richard L. Scheaffer for their genuine interest, incessant efforts and unlimited energy which made it possible to remove the stumbling stone out of my way to join the University of Florida. I would also like to express my gratitude to my respected teachers, especially to Professors Tathagata Bandyopadhyay, Prasanta Kumar Giri and Parthasarathi Lahiri for their encouragement and help which

were crucial to my coming to the United States. I feel very fortunate in being able to turn to them whenever necessary. I would also like to acknowledge the help and support I received from Krishnandu Ghosh in preparing me to come to the United States.

I am also grateful and highly indebted to my Alma Mater RamaKrishna Mission Residential College, Narendrapur, West Bengal, India, for all the support I received. Had I not been admitted to Narendrapur, I would have never pursued my studies in statistics. In this respect, I will always remember our Principal, Respected Swami Suparnananda Maharaj, and our Head of the Department, Dr. P.K. Giri, for their care and concern about me. I would like to offer my humble regards to them. I will take this opportunity to express my appreciation to Professor Ashok Kumar Hazra for the initiative he took in introducing me to Narendrapur. It is a great pleasure to acknowledge Professor Uttam Bandyopadhyay to whom I definitely owe a lot for my basic understanding of statistics as an undergraduate. The interest was further stimulated by the insightful teaching of Professor S.K. Chatterjee of Calcutta University when I was a master's student.

I would like to thank my parents for all they have done for me. I am deeply indebted to my numerous well-wishers and friends without whose generous support it would have been impossible for me to pursue my higher studies. I would like to express my sincere gratitude especially to

Mr. Bibekananda Nandi, Professor K.M. Senapati, Mrs. Durga Senapati, Mr. K.C. Ghosh and Mrs. Bimala Ghosh, who have always considered me as a part of their family. I also consider myself extremely lucky to get the affectionate concern of Mrs. Senapati, who has always treated me like her own son. My heartfelt thanks are for my "unofficial" host family in Gainesville, Dr. Malay Ghosh and his wife, our beloved Doladi.

I would like to thank Dr. A.P. Reznick of the United States Census Bureau for providing me with the computing facilities during my stay in the Bureau as an ASA/NSF Research Associate. Last but not least, I would like to thank Ms. Cindy Zimmerman for her skillful typing in putting a scribbled manuscript into final form.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	iii
ABSTRACT	viii
CHAPTERS	
ONE INTRODUCTION	1
1.1 Literature Review	1
1.2 The Subject of This Dissertation	10
TWO BAYESIAN PREDICTION OF MEANS IN LINEAR MODELS: GENERAL CASE	15
2.1 Introduction	15
2.2 Description of the Hierarchical Bayes Model with Examples	18
2.3 Hierarchical Bayes Analysis	31
2.4 Applications of Hierarchical Bayes Analysis	37
2.5 Hierarchical Bayes Prediction of Finite Population Mean Vector in Absence of Unit Level Observations	57
THREE OPTIMALITY OF BAYES PREDICTORS FOR MEANS IN A SPECIAL CASE	65
3.1 Introduction	65
3.2 The Hierarchical Bayes Predictor in a Special Case	67
3.3 Best Unbiased Prediction and Stochastic Domination in Small Area Estimation	72
3.4 Best Unbiased Prediction and Stochastic Domination in Infinite Population	88
3.5 Best Equivariant Prediction in Small Area Estimation	92
3.6 Best Equivariant Prediction in Infinite Population	109

FOUR	ASYMPTOTIC OPTIMALITY OF HIERARCHICAL BAYES PREDICTORS FOR MEANS	114
4.1	Introduction	114
4.2	Model, Loss, Prior and Predictors	117
4.2.1	General Expressions for Bayes Risks Difference.....	119
4.2.2	Random Regression Coefficients Model.....	120
4.2.3	Nested Error Regression Model.....	129
4.3	Asymptotic Optimality with Known First Stage Variance Component: Fay-Herriot Model	137
4.4	Asymptotic Optimality with Unknown Variance Components	153
FIVE	SIMULTANEOUS BAYESIAN ESTIMATION OF SMALL AREA VARIANCES	169
5.1	Introduction	169
5.2	Bayes Estimation of a Quadratic Form when Ratios of Variance Components Known	171
5.3	Asymptotic Optimality in Nested Error Regression Model for Known Ratio of Variance Components	173
5.4	Asymptotic Optimality in Nested Error Regression Model with Unknown Variance Components	182
SIX	SUMMARY AND FUTURE RESEARCH	189
6.1	Summary	189
6.2	Future Research	190
APPENDICES		
A	PROOF OF THEOREM 2.3.1	194
B	AN INDEPENDENCE RESULT IN A FAMILY OF ELLIPTICALLY SYMMETRIC DISTRIBUTIONS	203
BIBLIOGRAPHY		207
BIOGRAPHICAL SKETCH		213

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

BAYESIAN PREDICTION IN MIXED LINEAR MODELS
WITH APPLICATIONS IN SMALL AREA ESTIMATION

BY

GAURI SANKAR DATTA

August, 1990

Chairman: Dr. Malay Ghosh
Major Department: Statistics

Small area estimation is gaining increasing popularity in recent times. Government agencies in the United States and Canada have been involved in estimating unemployment rates, per capita income, crop yield, etc. simultaneously for many state and local government regions. Typically, only a few samples are available from an individual area. Consequently, reliable estimators of "parameters," such as the mean or the variance for the area, need to "borrow strength" from similar neighboring areas implicitly or explicitly through a model. Such estimators usually have a smaller mean squared error of prediction than the survey estimators.

In this dissertation, a general hierarchical Bayes (HB) model is considered for small area estimation. Some of the widely used models in small area estimation including the nested error regression model, random

regression coefficients model, etc. considered by earlier authors are seen to be special cases of the proposed general model. The predictive distribution of a characteristic of interest for the unsampled population units is found given the observations on the sampled units and is used to draw inference. In particular, simultaneous estimators of several small area means and variances are developed. A mixed linear model with noninformative prior for regression coefficients (or fixed effects) and independent gamma priors (possibly noninformative) for the inverse of the variance components is used.

In a special case of this HB analysis, when the vector of the ratios of the variance components is known, the HB predictor of the vector of means in finite population sampling is shown to possess some frequentist optimal properties (such as best unbiased predictor, best equivariant predictor, etc.) basically under the elliptical symmetry assumptions.

Performance of this HB predictor is evaluated by comparing its Bayes risk with that of subjective Bayes predictor with "true" or "elicited" prior for the unknown superpopulation parameters. It is shown that, under a balanced one-way random effects model with covariates and average squared error loss, the difference in the Bayes risks of the HB predictor and the "true" Bayes predictor of the finite population mean or variance vector approaches zero as the number of small areas becomes increasingly large.

CHAPTER ONE INTRODUCTION

1.1 Literature Review

Use of linear models by astronomers for predicting the positions of celestial bodies goes back several centuries. Starting from these days, the use of model-based inference for prediction has received considerable attention. In particular, animal and plant breeders have used such models for predicting some characteristics of the future progeny. Starting with the pioneering work of Henderson (1953), considerable attention has been devoted to this problem. We refer to Gianola and Fernando (1986) and Harville (in press) where other references are cited. On the other hand, survey analysts have used the model-based approach in finite population sampling with the goal of predicting certain characteristics of the unsampled units in the population on the basis of the observed sample. Early work on this topic may be found in Cochran (1939, 1946) where the finite population is viewed as a realization from a hypothetical superpopulation.

In recent years, small area (domain) estimation has grown into an important topic in survey sampling. Use of

small area statistics was in existence as early as the 11th century in England and the 17th century in Canada (see Brackstone, 1987). However, these early small area statistics were based on data obtained by complete enumeration.

Due to the availability of limited resources and the advent of sophisticated statistical methodologies, for the past few decades, sample surveys, for most purposes, have been widely used as the means of data collection in contrast to complete enumeration. The data collected from these surveys have been very effectively used to provide suitable statistics at the national and state levels on a regular basis. However, the use of survey data in sublevels below the state level (for example, county or other subdivision) was limited because the estimates for these small areas usually were based on small samples and produced unacceptably large standard errors and coefficients of variation. To improve the reliability of the small area statistics, it is necessary to have a much larger sample size for an individual area than can be afforded with the limited resources available. Consequently, the use of survey data (possibly in association with the census data) in producing reliable small area statistics did not receive much attention.

During the last few years, many countries, including the United States and Canada, have recognized the importance of small area estimation. Recently, there is a growing concern among several governments with the issues of distribution, equity and disparity. There may exist subgroups within a given population which are far below the average in certain respects, thereby necessitating remedial action on the part of the government. Before taking such an action, there is a need to identify such subgroups, and accordingly, the statistical data at the relevant subgroup levels must be available. So, different government agencies like the Census Bureau, Bureau of Labor Statistics, Statistics Canada and Central Bureau of Statistics of Norway have been involved in obtaining estimates of population counts, adjustment factors to census counts, unemployment rates, per capita income, etc. for state and local government areas.

In the face of this problem, small area estimation techniques have emerged that "borrow strength" from similar neighboring areas for estimation and prediction purposes. Through use of some appropriate model and auxiliary information (possibly obtained through complete enumeration, for example, census or satellite), small area estimators of the parameters of interest (such as the finite population mean, variance, etc.) usually improve

over the survey estimators. For a good review of the small area estimation literature one may refer to Ghosh and Rao (1990).

The necessity of "borrowing strength" has been realized by many statisticians. Ericksen (1974) advocated the use of regression method for estimating population changes of local areas. Fay and Herriot (1979) proposed an adaptation of the James-Stein estimator to survey estimates of income for small areas. Survey estimates being based on a small sample size (which is usually 20 percent of population of size less than 1000) usually have large standard errors and coefficients of variation. To rectify this, these authors first fit a regression equation to the census sample estimates, using as independent variables the county values, tax return data for the year 1969 and data for housing from the 1970 census. The estimate they provided for each place was a weighted average of the sample estimate and the regression estimate. Battese, Harter and Fuller (1988) considered prediction of areas under corn and soybeans for 12 counties in north-central Iowa based on 1978 June Enumerative Survey and LANDSAT satellite data. Battese, Harter and Fuller (BHF) used a linear regression model defining a relationship between the survey and satellite data and used this relationship to obtain predictors of mean crop areas per segment in the

sampled counties. Fuller and Harter (1987) also considered a multivariate extension of this model.

There is a similar problem of prediction faced by the animal breeders. For the purpose of selecting the best animals for future breeding, they need to come up with an index for each animal under consideration. Henderson (1953, 1975) advocated the use of best linear unbiased predictor (BLUP) of certain linear combinations of fixed and random effects using a mixed linear model. Harville (in press) used a mixed linear model for predicting the average weight of single-birth male lambs which are progeny of sires belonging to different population lines and dams belonging to different age categories. Harville and Fenech (1985) considered this example for estimating the heritability.

Yet other problems based on a linear model come up in varietal trials and comparative experiments. In comparative experiments, several treatments have to be compared and their effects or some suitable contrasts have to be estimated. Multicentered clinical trials are good examples of comparative experiments (see Fleiss, 1986). Problems of this type and the ones mentioned in the previous paragraph can be viewed as arising in an infinite population framework as opposed to the problems in finite population sampling.

The methods that have usually been proposed in model-based inference use either a variance components approach or an empirical Bayes (EB) approach, although as pointed out by Harville (1988, in press), the distinction between the two is often superfluous. Both these procedures use certain mixed linear models for prediction purposes. First, assuming the variance components to be known, certain BLUPs or EB predictors are obtained for the unknown parameters of interest. Then the unknown variance components are estimated typically by Henderson's method of fitting of constants or the restricted maximum likelihood (REML) method. The resulting estimators, which can be called estimated BLUPs or EBLUPs (see Harville, 1977), are used for final prediction purposes.

Empirical Bayes approach in small area estimation was first given in Fay and Herriot (1979), and later also used by Ghosh and Meeden (1986), Ghosh and Lahiri (1987a, 1988) among others. According to this procedure, first a Bayes estimate of the unknown parameter of interest is obtained by using a normal prior or using a linear Bayes argument (Hartigan, 1969). The unknown parameters of the prior are then estimated by some classical methods like the method of moments, method of maximum likelihood or some combination thereof. The resulting estimator of the parameter of interest is the so-called EB estimator.

Although the above approach of EBLUP or EB is usually quite satisfactory for point prediction, it is very difficult to estimate the standard errors associated with these predictors. This is primarily due to the lack of closed form expressions for the mean squared errors (MSEs) of the EBLUPs or the EB predictors. Kackar and Harville (1984) suggested an approximation to the MSEs (also Harville, 1985, 1988, in press; Harville and Jeske, 1989). Prasad and Rao (1990) proposed estimates of these approximate MSEs in three specific mixed linear models. All these approximations rest heavily on the normality assumption. Recently, Lahiri and Rao (1990) considered this problem, relaxing the normality assumption, assuming some moment conditions without the presence of auxiliary information. The work of Prasad and Rao (1990) suggests that their approximations work well when the number of small areas is sufficiently large. It is not clear though how these approximations fare for a small or even moderately large number of small areas.

Ghosh and Lahiri (in press) proposed an HB procedure as an alternative to the EBLUP or the EB procedure. In an HB procedure, if one uses the posterior mean for estimating the parameter of interest, then a natural estimate of the standard error associated with this estimator is its posterior standard deviation (s.d.). The estimate, though

often complicated, can be found exactly via numerical integration without approximation.

The model considered by Ghosh and Lahiri (in press) was, however, only a special case of the so-called nested error regression model, also used by BHF. A similar model was considered by Stroud (1987), but his general analysis was performed only for the balanced case, that is when the number of samples was the same for each stratum.

Other models have also been proposed. In a recent article, Choudhry and Rao (1988) considered five specific models for small area estimation not included in the earlier work of Prasad and Rao (1990). Recently, Royall (1979) and Lui and Cumberland (1989) considered certain cross-classificatory models for small area estimation. The latter carried out a Bayesian analysis assuming the degeneracy of certain terms in an usual two-way linear model.

For a Bayesian analysis in the context of animal breeding, one may refer to Gianola and Fernando (1986). However, they did not consider the HB analysis. They used subjective informative priors which are constructed from the previous data and experiments. Also, they showed how some of the classical measures in animal breeding can have Bayesian justification. They have also considered noninformative improper prior.

One important special case which arises in the above approaches which is also important in the theory of least squares. When the ratios of variance components are known, the predictors (least squares, empirical Bayes or hierarchical Bayes) are BLUPs (Henderson, 1963). For related BLUP results for predicting scalars in finite population sampling one may refer to Royall (1979), Lui and Cumberland (1989), Prasad and Rao (1990) and several others. Harville (1985, 1988, in press) has pointed out the BLUP properties of Bayesian scalars in general mixed linear models (see also Harville, 1976). Ghosh and Lahiri (in press) have extended Henderson and others scalar BLUP notion to show the Bayesian predictor of the vector of finite population mean is BLUP.

To conclude this discussion, we will briefly mention another problem. So far, we have considered only the problem of estimating the mean in finite population sampling. Another important problem in finite population sampling is estimating the finite population variance. Ericson (1969) found the Bayes estimator of finite population variance under a normal theory set up. Empirical Bayes estimation of finite population variance in small area estimation was considered by Ghosh and Lahiri (1987b) and Lahiri and Tiwari (in press) without the presence of auxiliary information.

1.2 The Subject of This Dissertation

In this dissertation, we present a unified Bayesian prediction theory for linear models in small area estimation in the context of finite population sampling. A general Bayesian model is presented which can be regarded as an extension of the HB ideas of Lindley and Smith (1972) to prediction. This general model can also be applied in infinite population situations, for example, in animal breeding and in other applications where a mixed linear model is used.

In Chapter Two, we introduce a general HB model and use this model for simultaneous estimation of several small area means in finite population sampling. Some of the widely used models in small area estimation including the nested error regression model (Battese et al., 1988; Prasad and Rao, 1990; Stroud, 1987; Ghosh and Lahiri, in press), the random regression coefficients model (Dempster et al., 1981; Prasad and Rao, 1990), the cross-classificatory models (Royall, 1979; Lui and Cumberland, 1989) and multi-stage sampling models (Ghosh and Lahiri, 1988; Malec and Sedransk, 1985; Scott and Smith, 1969) can be regarded as special cases of our model. The posterior distribution as well as the resulting posterior means and variances of the unobserved units in the population given the sample units are provided in this chapter. Also for an infinite

population, given the data, the conditional distribution and the conditional mean and variance of the vector of effects are provided. These two analyses are applied to two real data sets. It is worthwhile to mention that Bayesian analysis in linear models was initiated by Hill (1965). See also Hill (1977, 1980). For a good exposition on HB analysis see Berger (1985).

In Chapter Three, a special case of HB models discussed in the previous chapter is considered. Based on this model, which assumes known ratios of variance components, certain optimal properties of the HB predictors proposed in this chapter are proved. Although, developed within a Bayesian framework, these results should be of appeal also to frequentists. The BLUP notion for real valued parameters is extended to vector valued parameters, and it is shown that the Bayesian predictors derived in this chapter are indeed BLUPs. From this, as a special case, it follows that the Bayesian predictors of the finite population mean vector and other linear parameters are BLUPs as well. Our BLUP result for the finite population mean vector unifies a number of similar results derived under specific models (e.g., Royall, 1979; Ghosh and Lahiri, in press; Lui and Cumberland, 1989; Prasad and Rao, 1990). Like other related articles, our BLUP results do not require any normality assumption of the model. For a

suitable subclass of elliptically symmetric distributions, including but not limited to the normal, the HB predictors are shown to be best unbiased; that is, they have the smallest variance-covariance matrix within the class of all unbiased predictors. Also, following Hwang (1985), we have been able to show that the BLUPs also “universally” (or “stochastically”) dominate the linear unbiased predictors for elliptically symmetric distributions. The notion of “universal” and “stochastic” domination will be made precise in Chapter Three. Also, it is established that under a suitable group of transformations, the HB predictors are best within the class of all equivariant predictors for elliptically symmetric distributions. Jeske and Harville (1987) have shown that the scalar BLUPs are best equivariant within the class of all linear equivariant predictors without any distributional assumption. However, to our knowledge, the equivariance results for vector valued predictors have not been addressed before in this context in their full generality.

In Chapter Four, we have established some asymptotic results regarding the Bayes risk performance of certain HB predictors of the finite population mean vector. We have considered two specific models, namely, the random regression coefficients model and the nested error regression model, introduced in Chapter Two. We have shown

that under average squared error loss the Bayes risk difference between the HB predictors and the subjective Bayes predictors for a "true" prior goes to zero as the number of small areas goes to infinity. This shows our HB predictors are "asymptotically optimal" (A.O.) in the sense of Robbins (1955). The A.O. property of certain EB predictors arising naturally in the context of finite population sampling was proved in Ghosh and Meeden (1986), Ghosh and Lahiri (1987a) and Ghosh, Lahiri and Tiwari (1989).

Chapter Five is devoted to the simultaneous estimation of several strata variances. We have considered the special cases of the nested error regression model as considered by Ghosh and Lahiri (in press) and Stroud (1987) in detail. As in Chapter Four, we have proved the A.O. property of these predictors. Ghosh and Lahiri (1987b) and Lahiri and Tiwari (in press) have proved the A.O. property of certain EB predictors of finite population variances.

We reemphasize that the present dissertation provides a unified Bayesian analysis both in the finite and infinite population framework. For finite population, we unify a number of models considered earlier by different authors. From the analysis of two data sets we undertook in Chapter Two, it is clear that the proposed procedure is a viable alternative to the available EBLUP or EB procedures. Also,

to our knowledge, estimates of MSEs or good approximations thereof are not available except for a few specific models. The Bayesian procedures of this dissertation, on the other hand, can serve as a general recipe to handle a greater variety of problems. Also, the inferential methods of the following chapters are implementable for data analysis, especially in these days of sophisticated computing facilities.

CHAPTER TWO
BAYESIAN PREDICTION OF MEANS IN LINEAR MODELS:
GENERAL CASE

2.1 Introduction

In this chapter we will consider two similar but different problems simultaneously. One problem refers to the small area estimation problem in the context of finite population sampling and the other problem deals with the prediction problem in comparative experiments in the context of ANOVA, ANOCOVA or linear regression in infinite population situation. In both these cases, a mixed linear model is used. In the first case, we are interested in predicting some finite population characteristics (e.g., finite population totals or means) whereas in the second case we are interested in predicting linear functions of fixed and random effects.

In the finite population sampling set up, we assume that there are m strata, the i^{th} stratum U_i containing a finite number of units N_i with units labelled $U_{i1}, \dots, U_{iN_i}$. Let Y_{ij} denote some characteristic of interest associated with the j^{th} unit in the i^{th} stratum ($j = 1, \dots, N_i$; $i = 1, \dots, m$). We assume that the Y_{ij} are observables in the sense that we can actually learn the exact value of any

of the Y_{ij} for some finite cost. We are interested in predicting some linear combinations of these observables (like the finite population total or mean for each small area or domain) using a quadratic loss. For notational convenience, we will denote a sample of size n_i from the i^{th} stratum by $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$.

On the other hand in the infinite population set up we are interested in predicting linear combinations (in particular contrasts) of fixed and random effects. Note that in this set up these quantities are not observables. For this problem too, we use a quadratic loss function.

We will use the word predictands to refer to the quantities we want to predict in both the problems. The analysis will be done in two stages; in the first stage we assume the ratios of the variance components are known whereas in the second stage we consider the more general situation where all the variance components are unknown.

In Section 2.2 a general HB model will be described and a number of interesting examples arising in finite population sampling or in the infinite population set up will be considered. Some of the existing models used in the context of finite population sampling are shown to be special cases of the general model to be proposed in this section.

Realizing the importance of the problem, the most general situation where all the variance components are unknown will be considered in this chapter, whereas the known ratios of variance components will be considered in the next chapter. This general situation is considered in Section 2.3. A general mixed linear model is considered and some prior distribution is assigned to all unknown parameters which consist of the vector of fixed effects and the variance components. In the first part of Section 2.3, for the model introduced in Section 2.2, we have found the posterior (predictive) distribution of the characteristic of interest of the nonsampled population units given the values of that characteristic for the sample units in finite population sampling. Also the posterior mean vector and posterior variance-covariance matrix corresponding to the characteristic vector of nonsampled units are obtained from this predictive distribution. In particular, the posterior means and the variances of the finite population means for all the small areas are obtained. In the second half of Section 2.3, we have obtained the posterior distribution of the vector of fixed and random effects for the model introduced in Section 2.2. In particular, expressions for the posterior means and variances of certain predictands are developed.

In Section 2.4, we have applied the results of Section 2.3 to some actual data sets. First, we shall consider the corn and soybeans data which appeared in Battese, Harter and Fuller (1988). Using the HB analysis developed in Section 2.3, we have derived the posterior means and posterior standard deviations for the 12 small area (county) means. The second data set containing the weights of 62 single-birth lambs appeared in Harville (in press). This is analyzed by HB methods for an infinite population set up developed in the second half of Section 2.3.

Finally, in Section 2.5 an HB analysis of the model considered by Carter and Rolph (1974) and subsequently by Fay and Herriot (1979) to estimate the per capita income of small places is considered. In this situation unit level observations are not available and we are interested in predicting the finite population mean for each small area. Here the sampling variances are different and are assumed to be known; also, we put a uniform prior on the regression coefficients and a gamma prior (proper or improper) on the inverse of the prior variance.

2.2 Description of the Hierarchical Bayes Model with Examples

Consider the following Bayesian model:

(A) conditional on $\mathbf{b} = (b_1, \dots, b_p)^T$, $\mathbf{v} = (v_1, \dots, v_q)^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_t)^T$ and r , let

$$\underline{Y} \sim N(\underline{X}\underline{b} + \underline{Z}\underline{v}, r^{-1}\underline{\Psi});$$

(B) conditional on $\underline{b}$, $\underline{\lambda}$ and r , let

$$\underline{v} \sim N(\underline{0}, r^{-1}\underline{D}(\underline{\lambda}));$$

(C) $\underline{B}$, $\underline{R}$ and $\underline{\Lambda}$ have a certain joint prior distribution proper or improper.

Stages (A) and (B) of the model can be identified as a general mixed linear model. To see this, write

$$\underline{Y} = \underline{X}\underline{b} + \underline{Z}\underline{v} + \underline{e}, \quad (2.2.1)$$

where $\underline{b}$ is the vector of fixed effects, $\underline{e}$ and $\underline{v}$ are mutually independent with $\underline{e} \sim N(\underline{0}, r^{-1}\underline{\Psi})$ and $\underline{v} \sim N(\underline{0}, r^{-1}\underline{D}(\underline{\lambda}))$; $\underline{X}$ and $\underline{Z}$ are known design matrices, $\underline{\Psi}$ is a known positive definite (p.d.) matrix, while $\underline{D}(\underline{\lambda})$ is a p.d. matrix which is structurally known except possibly for some unknown $\underline{\lambda}$. In the examples to follow, $\underline{\lambda}$ involves the ratios of variance components.

In the context of small area estimation, partition $\underline{Y}(N_T \times 1)$, $\underline{X}(N_T \times p)$, $\underline{Z}(N_T \times q)$ and $\underline{e}(N_T \times 1)$ with conformity, and rewrite the model given in (2.2.1) as

$$\begin{pmatrix} \underline{Y}^{(1)} \\ \underline{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix} \underline{b} + \begin{pmatrix} \underline{Z}^{(1)} \\ \underline{Z}^{(2)} \end{pmatrix} \underline{v} + \begin{pmatrix} \underline{e}^{(1)} \\ \underline{e}^{(2)} \end{pmatrix}. \quad (2.2.2)$$

In the above paragraph, $\underline{Y}^{(1)}(n_T \times 1)$ corresponds to the vector of sampled units from m small areas or strata, while

$Y^{(2)}((N_T - n_T) \times 1)$ corresponds to the vector of unsampled units. We will further partition $Y^{(1)T}$ into $Y^{(1)T} = (Y_1^{(1)T}, \dots, Y_m^{(1)T})$, where the n_i -component vector $Y_i^{(1)} = (Y_{i1}, \dots, Y_{in_i})^T$ is the vector of sampled units from the i^{th} small area. Similarly, $Y^{(2)T}$ can be partitioned into $(Y_1^{(2)T}, \dots, Y_m^{(2)T})$, where the $(N_i - n_i)$ -component vector $Y_i^{(2)} = (Y_{i, n_i+1}, \dots, Y_{i, N_i})^T$ is the vector of unsampled units for the i^{th} small area. One of our primary objectives in small area estimation is to estimate the finite population mean vector $\gamma = (\gamma_1, \dots, \gamma_m)^T$ where $\gamma_i = \sum_{j=1}^{N_i} Y_{ij} / N_i$, $i = 1, \dots, m$. More generally, we may be interested in predicting the vector of linear combinations $A Y^{(1)} + C Y^{(2)} = \xi(Y^{(1)}, Y^{(2)})$ (say) for known matrices $A(u \times n_T)$ and $C(u \times (N_T - n_T))$. For this purpose it suffices to find the predictive distribution of $Y^{(2)}$ given $Y^{(1)}$. In the next section this will be accomplished by using model-based approach in survey sampling.

Before we consider the other problem in the infinite population set up, we identify some of the existing models introduced for small area estimation by several authors as special cases of (2.2.2). In what follows, we shall use the notation I_u for a $u \times u$ identity matrix, $\mathbf{1}_u$ for a u -component column vector with each element equals to 1, $J_{u,v}$ for the $u \times v$ matrix $\mathbf{1}_u \mathbf{1}_v^T$ and J_u for the $u \times u$ matrix $J_{u,u}$.

Also, let $_{1 \leq i \leq k}^{col} (B_i)$ denote the matrix $(B_1^T, \dots, B_k^T)^T$, and let

$$_{i=1}^k \oplus A_i \text{ denote the matrix } \begin{bmatrix} A_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & A_k \end{bmatrix}.$$

First, consider the nested error regression model

$$Y_{ij} = x_{ij}^T b + v_i + e_{ij} \quad (j = 1, \dots, N_i; i = 1, \dots, m). \quad (2.2.3)$$

The model was considered by Battese, Harter and Fuller (1988). They assumed the v_i and e_{ij} to be mutually independent with v_i iid $N(0, (\lambda r)^{-1})$, and e_{ij} iid

$N(0, r^{-1})$. In this case $\underline{x}^{(1)} = _{1 \leq i \leq m}^{col} \left[_{1 \leq j \leq n_i}^{col} x_{ij}^T \right]$,

$$\underline{x}^{(2)} = _{1 \leq i \leq m}^{col} \left[_{n_i+1 \leq j \leq N_i}^{col} x_{ij}^T \right], \quad \underline{z}^{(1)} = \bigoplus_{i=1}^m 1_{n_i}, \quad \underline{z}^{(2)} = \bigoplus_{i=1}^m 1_{N_i - n_i},$$

$\underline{\Psi} = I_{N_T}$, $t = 1$, $\underline{\lambda} = \lambda$ and $\underline{D}(\underline{\lambda}) = \lambda^{-1} I_m$. In the further special case of Ghosh and Lahiri (in press), $x_{ij} = x_i$ for every $j = 1, \dots, N_i$, $i = 1, \dots, m$. Note that $\lambda = V(e_{ij})/V(v_i)$, a ratio of the variance components.

The random regression coefficients model of Dempster, Rubin and Tsutakawa (1981) (also Prasad and Rao, 1990) is also a special case of ours. In this set up, $\underline{x}^{(1)}$, $\underline{x}^{(2)}$, $\underline{\Psi}$ and $\underline{D}(\underline{\lambda})$ are the same as in the nested error regression

model, but $\underline{z}^{(1)} = \bigoplus_{i=1}^m \left[_{1 \leq j \leq n_i}^{col} x_{ij}^T \right]$ and $\underline{z}^{(2)} = \bigoplus_{i=1}^m \left[_{n_i+1 \leq j \leq N_i}^{col} x_{ij}^T \right]$.

Some of the models of Choudhry and Rao (1988) can also be treated as special cases of ours. For example, one of their models is given by

$$Y_{ij} = bx_{ij} + v_i + e_{ij}^* x_{ij}^{1/2} \quad (j=1, \dots, N_i; i=1, \dots, m) \quad (2.2.4)$$

with v_i iid $N(0, (r\lambda)^{-1})$ and e_{ij}^* iid $N(0, r^{-1})$. Here $p = 1$, $b_1 = b$, $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are the same as in the nested error regression model with the vector $\underline{x}_{ij}$ replaced by scalar x_{ij} ; $\underline{z}^{(1)}$, $\underline{z}^{(2)}$ and $\underline{D}(\lambda)$ are the same as in the nested error regression model, but $\underline{\Psi} = \text{Diag}(x_{11}, \dots, x_{1N_1}, \dots, x_{m1}, \dots, x_{mN_m})$. Another model considered by these authors is similar to the one given in (2.2.4) with x_{ij} replacing $x_{ij}^{1/2}$ as multipliers of e_{ij}^* . Yet another model considered by Choudhry and Rao (1988) is

$$Y_{ij} = bx_{ij} + v_i x_{ij}^{1/2} + e_{ij}^* x_{ij}^{1/2} \quad (j = 1, \dots, N_i; i = 1, \dots, m), \quad (2.2.5)$$

with the v_i and e_{ij}^* having the same distribution as in the previous model. Here $\underline{\Psi} = \text{Diag}(x_{11}, \dots, x_{1N_1}, \dots, x_{m1}, \dots, x_{mN_m})$, $\underline{z}^{(1)} = \bigoplus_{i=1}^m \underline{u}_i^{(1)}$, $\underline{z}^{(2)} = \bigoplus_{i=1}^m \underline{u}_i^{(2)}$, with $\underline{u}_i^{(1)} = (x_{i1}^{1/2}, \dots, x_{iN_i}^{1/2})^T$, $\underline{u}_i^{(2)} = (x_{i, n_i+1}^{1/2}, \dots, x_{iN_i}^{1/2})^T$; $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are the same as in (2.2.4).

It is possible also to include certain cross-classification models as special cases of our general

linear model. For example, suppose there are m small areas labelled $1, \dots, m$. Within each small area, units are further classified into c subgroups (socioeconomic class, age, etc.) labelled $1, \dots, c$. The cell sizes N_{ij} ($i = 1, \dots, m; j = 1, \dots, c$) assumed to be known. Let Y_{ijk} ($k = 1, \dots, N_{ij}$) denote the measurement on the k^{th} individual in the $(i, j)^{\text{th}}$ cell. Conditional on b , r and λ , suppose

$$Y_{ijk} = x_{ij}^T b + \tau_i + \eta_j + \gamma_{ij} + e_{ijk} \quad (2.2.6)$$

($k = 1, \dots, N_{ij}; i = 1, \dots, m; j = 1, \dots, c$), with τ_i , η_j , γ_{ij} and e_{ijk} mutually independent with $e_{ijk} \text{ iid } N(0, r^{-1})$, $\gamma_{ij} \text{ iid } N(0, (\lambda_3 r)^{-1})$, $\eta_j \text{ iid } N(0, (\lambda_2 r)^{-1})$ and $\tau_i \text{ iid } N(0, (\lambda_1 r)^{-1})$. In this case

$$Y^{(1)} = {}_{1 \leq i \leq m} \text{col} \left({}_{1 \leq j \leq c} \text{col} \left({}_{1 \leq k \leq n_{ij}} Y_{ijk} \right) \right),$$

$$Y^{(2)} = {}_{1 \leq i \leq m} \text{col} \left({}_{1 \leq j \leq c} \text{col} \left({}_{n_{ij}+1 \leq k \leq N_{ij}} Y_{ijk} \right) \right),$$

$$x^{(1)} = {}_{1 \leq i \leq m} \text{col} \left\{ {}_{1 \leq j \leq c} \text{col} \left({}_{1 \leq k \leq n_{ij}} x_{ijk}^T \right) \right\},$$

$$x^{(2)} = {}_{1 \leq i \leq m} \text{col} \left\{ {}_{1 \leq j \leq c} \text{col} \left({}_{1 \leq k \leq N_{ij}-n_{ij}} x_{ijk}^T \right) \right\},$$

$$Z^{(1)} = \left(Z_1^{(1)} \ Z_2^{(1)} \ Z_3^{(1)} \right) \text{ where } Z_1^{(1)} = \bigoplus_{i=1}^m 1_{n_{i.}},$$

$$n_{i.} = \sum_{j=1}^c n_{ij},$$

$$\mathbb{Z}_2^{(1)} = \underset{1 \leq i \leq m}{\text{col}} \left\{ \underset{j=1}{\overset{c}{\oplus}} \mathbb{1}_{n_{ij}} \right\}, \quad \mathbb{Z}_3^{(1)} = \underset{i=1}{\overset{m}{\oplus}} \left\{ \underset{j=1}{\overset{c}{\oplus}} \mathbb{1}_{n_{ij}} \right\}$$

and $\mathbb{Z}^{(2)}$ is a matrix similar to $\mathbb{Z}^{(1)}$ with $(N_{ij} - n_{ij})$ replacing n_{ij} in defining the dimensions of the vectors. Also, $\mathbf{v}^T = (\tau_1, \dots, \tau_m, \eta_1, \dots, \eta_c, \gamma_{11}, \dots, \gamma_{mc})$, $t = 3$, $\lambda^T = (\lambda_1, \lambda_2, \lambda_3)$ and $\mathbb{D}(\lambda) = \text{Diag}(\lambda_1^{-1} \mathbb{I}_m, \lambda_2^{-1} \mathbb{I}_c, \lambda_3^{-1} \mathbb{I}_{mc})$. Special cases of this model have been considered by several others. Lui and Cumberland (1989) considered a model where τ_i and γ_{ij} are degenerate at zeroes. Also, they assumed the variance ratio λ_2 to be known in deriving their estimators, and did not address the issue of unknown λ_2 appropriately.

Next we show that the two stage sampling model with covariates and m strata is a special case of our general linear model. Suppose that the i^{th} stratum contains L_i primary units. Suppose also that the j^{th} primary unit within the i^{th} stratum contains N_{ij} subunits. Let Y_{ijk} denote the value of the characteristic of interest for the k^{th} subunit within the j^{th} primary unit from the i^{th} stratum ($k = 1, \dots, N_{ij}$; $j = 1, \dots, L_i$; $i = 1, \dots, m$). From the i^{th} stratum, a sample of ℓ_i primary units is taken. For the j^{th} selected primary unit within the i^{th} stratum, a sample of n_{ij} subunits are selected. For notational convenience, denote, without loss of generality, the sample values by Y_{ijk} ($k = 1, \dots, n_{ij}$; $j = 1, \dots, \ell_i$; $i = 1, \dots, m$).

Assume conditional on $\mathbf{b}$, r and λ

$$Y_{ijk} = \mathbf{x}_{ij}^T \mathbf{b} + \xi_i + \eta_{ij} + e_{ijk}$$

$$(k = 1, \dots, N_{ij}; j = 1, \dots, L_i; i = 1, \dots, m), \quad (2.2.7)$$

where ξ_i , η_{ij} and e_{ijk} are mutually independent with ξ_i iid $N(0, (\lambda_1 r)^{-1})$, η_{ij} iid $N(0, (\lambda_2 r)^{-1})$, e_{ijk} iid $N(0, r^{-1})$.

Let

$$\mathbf{Y}^{(1)} = \underset{1 \leq i \leq m}{\text{col}} \left[\underset{1 \leq j \leq \ell_i}{\text{col}} \left\{ \underset{1 \leq k \leq n_{ij}}{\text{col}} (Y_{ijk}) \right\} \right],$$

$$\mathbf{Y}^{(2)} = \underset{1 \leq i \leq m}{\text{col}} \left[\underset{1 \leq j \leq L_i}{\text{col}} \left\{ \underset{u_{ij} \leq k \leq N_{ij}}{\text{col}} (Y_{ijk}) \right\} \right],$$

$$u_{ij} = 1 + n_{ij} I[j \leq \ell_i], \quad i = 1, \dots, m,$$

$$\mathbf{v} = \left(\mathbf{s}^T \quad \mathbf{w}_1^T \quad \mathbf{w}_2^T \right)^T, \quad \mathbf{s} = \underset{1 \leq i \leq m}{\text{col}} (\xi_i),$$

$$\mathbf{w}_1 = \underset{1 \leq i \leq m}{\text{col}} \left(\underset{1 \leq j \leq \ell_i}{\text{col}} (\eta_{ij}) \right)$$

$$\text{and } \mathbf{w}_2 = \underset{1 \leq i \leq m}{\text{col}} \left(\underset{\ell_i+1 \leq j \leq L_i}{\text{col}} (\eta_{ij}) \right).$$

Also, let $\mathbf{e}^{(i)}$ be defined similarly as $\mathbf{Y}^{(i)}$, $i = 1, 2$.

Then (2.2.7) can be written as (2.2.2) with

$$\mathbf{x}^{(1)} = \underset{1 \leq i \leq m}{\text{col}} \left\{ \underset{1 \leq j \leq \ell_i}{\text{col}} \left(\mathbf{1}_{n_{ij}} \mathbf{x}_{ij}^T \right) \right\},$$

$$\underline{X}^{(2)} = \left\{ \underset{1 \leq i \leq m}{\text{col}} \left\{ \underset{1 \leq j \leq L_i}{\text{col}} \left(\mathbf{1}_{r_{ij}} \mathbf{x}_{ij}^T \right) \right\} \right\}, \text{ where}$$

$$r_{ij} = N_{ij} - u_{ij} + 1;$$

$$\underline{Z}^{(1)} = \left(\underline{S}^{(1)} \quad \underline{W}^{(1)} \right), \quad \underline{S}^{(1)} = \bigoplus_{i=1}^m \mathbf{1}_{n_{i.}}, \quad n_{i.} = \sum_{j=1}^{\ell_i} n_{ij},$$

$$\underline{W}^{(1)} = \left(\underline{W}_1^{(1)} \quad \underline{W}_2^{(1)} \right) \text{ where } \underline{W}_1^{(1)} = \bigoplus_{i=1}^m \left(\bigoplus_{j=1}^{\ell_i} \mathbf{1}_{n_{ij}} \right),$$

$$\underline{W}_2^{(1)} = n_T \mathbf{0}_{L. - \ell.}, \quad n_T = \sum_{i=1}^m n_{i.}, \quad L. = \sum_{i=1}^m L_i, \quad \ell. = \sum_{i=1}^m \ell_i;$$

$$\underline{Z}^{(2)} = \left(\underline{S}^{(2)} \quad \underline{W}^{(2)} \right), \quad \underline{S}^{(2)} = \bigoplus_{i=1}^m \mathbf{1}_{R_i},$$

$$R_i = \sum_{j=1}^{L_i} N_{ij} - n_{i.} \text{ and } \underline{W}^{(2)} = \bigoplus_{i=1}^m \left(\bigoplus_{j=1}^{L_i} \mathbf{1}_{r_{ij}} \right).$$

Here $t = 2$, $\lambda = (\lambda_1, \lambda_2)^T$, $\Psi = I_{N_T}$, $D(\lambda) = \text{Diag}(\lambda_1^{-1} I_m, \lambda_2^{-1} I_{L.})$ with $N_T = \sum_{i=1}^m \sum_{j=1}^{L_i} N_{ij}$. The ideas can be extended directly to multistage sampling with more complicated notations. We may mention here that Bayesian analysis for two stage sampling was introduced first by Scott and Smith (1969) in a much simpler framework. A multistage analog of their work was provided by Malec and Sedransk (1985). Ghosh and Lahiri (1988) considered empirical Bayes estimation in multistage sampling.

Now we will consider the infinite population set up. In this context, we will use the model given by (A), (B) and (C) with the mixed linear model representation given by (2.2.1). Here we will assume the data vector $\underline{Y}$ is $n_T \times 1$ and the associated design matrices $\underline{X}$ and $\underline{Z}$ are $n_T \times p$ and $n_T \times q$ respectively. Without loss of generality, we also assume that $\text{rank}(\underline{X}) = p$. Our objective is to predict $\underline{S}\underline{b} + \underline{T}\underline{v} = \zeta(\underline{b}, \underline{v})$ (say) on the basis of $\underline{Y}$ where $\underline{S}(u \times p)$ and $\underline{T}(u \times q)$ are known matrices. Following the model-based inference, it suffices to find the posterior (conditional) distribution of $\begin{pmatrix} \underline{b} \\ \underline{v} \end{pmatrix} = \underline{W}$ (say) given $\underline{Y} = \underline{y}$. This conditional distribution is provided in the next section. We will conclude this section discussing a few specific models in the context of comparative trials and animal breeding which are special cases of the general model proposed in (2.2.1).

First consider multicentered clinical trial which is conducted in c participating clinics to compare two treatments, one already existing in the market and the other newly developed. Suppose there are n_{ij} subjects receiving the i^{th} treatment in the j^{th} clinic. Some of the n_{ij} could be zero. We are interested in estimating the treatments difference. For this example, we will use a mixed linear model for Y_{ijk} , some measured response corresponding to the k^{th} subject receiving the i^{th}

treatment in the j^{th} participating clinic. We consider the model:

$$Y_{ijk} = \mu_i + c_j + \gamma_{ij} + e_{ijk} \quad (2.2.8)$$

($k = 1, 2, \dots, n_{ij}$, $j = 1, \dots, c$, $i = 1, 2$) where c_j , γ_{ij} and e_{ijk} mutually independent with subject effects e_{ijk} are iid $N(0, r^{-1})$, treatment-clinic interaction γ_{ij} iid $N(0, (\lambda_2 r)^{-1})$ and clinic effects c_j iid $N(0, (\lambda_1 r)^{-1})$; μ_i is the effect due to the i^{th} treatment. Now we will write down the $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{b}$, $\mathbf{v}$ and $\mathbf{e}$ for (2.2.8). For ease of presentation, assume $n_{ij} > 0$ for all $i = 1, 2$, $j = 1, \dots, c$.

Then writing $\mathbf{Y} = (Y_{111}, \dots, Y_{11n_{11}}, \dots, Y_{1c1}, \dots, Y_{1cn_{1c}}, \dots, Y_{2c1}, \dots, Y_{2cn_{2c}})^T$, $\mathbf{b} = (\mu_1, \mu_2)^T$, $\mathbf{v} = (c_1, \dots, c_c, \gamma_{11}, \dots, \gamma_{1c}, \dots, \gamma_{2c})^T$, $\mathbf{X} = \bigoplus_{i=1}^2 \mathbf{1}_{n_i}$, $\mathbf{Z}_1 = \bigoplus_{j=1}^c \mathbf{1}_{n \cdot j}$, $\mathbf{Z}_2 = \bigoplus_{i=1}^2 \bigoplus_{j=1}^c \mathbf{1}_{n_{ij}}$, $\mathbf{Z} = (\mathbf{Z}_1 \ \mathbf{Z}_2)$, $\mathbf{e} = (e_{111}, \dots, e_{11n_{11}}, \dots, e_{1c1}, \dots, e_{1cn_{1c}}, \dots, e_{2c1}, \dots, e_{2cn_{2c}})^T$, $\mathbf{\Psi} = \mathbf{I}_{n_T}$, $\mathbf{\lambda} = (\lambda_1, \lambda_2)^T$ and $\mathbf{D}(\mathbf{\lambda}) = \text{Diag}(\lambda_1^{-1} \mathbf{I}_c, \lambda_2^{-1} \mathbf{I}_{2c})$, it is clear that (2.2.8) is a special case of (2.2.1) where in the above $n_{i \cdot} = \sum_{j=1}^c n_{ij}$, $i = 1, 2$, $n \cdot j = \sum_{i=1}^2 n_{ij}$, $j = 1, \dots, c$ and $n_T = \sum_{i=1}^2 \sum_{j=1}^c n_{ij}$ = total number of observations. We want to estimate $\mu_1 - \mu_2$ which is a special case of $\zeta(\mathbf{b}, \mathbf{v})$ with $\mathbf{S} = (1 \ -1)$ and $\mathbf{T} = \mathbf{0}^T$.

Next consider the animal breeding experiment. Suppose b breeds are randomly selected and d different diets are

used. Once again, one can use the linear model given by (2.2.8) for inferential purpose. In this case one may be interested in predicting some suitable linear functions of random quantities, c_j and γ_{ij} , known as the breeding values. These predicted breeding values can be used as a selection index for selecting the most suitable breeds for future breeding purpose.

As a concrete example in animal breeding we will discuss the example considered by Harville (in press) which involves prediction of the average birth weights of an infinite number of single-birth male lambs that are offspring of different sires in different population lines. The data consist of the weights (at birth) of 62 single-birth male lambs, and came from five distinct population lines. Each lamb was the progeny of one of 23 rams, and each lamb had a different dam. Age of the dam was recorded as belonging to one of three categories, numbered 1 (1-2 years), 2 (2-3 years), and 3 (over 3 years). Let Y_{ijkd} represent the weight (at birth) of the d^{th} of those lambs that are the offspring of the k^{th} sire in the j^{th} population line and a dam belonging to the i^{th} age category. Following Harville (in press), we will use the mixed linear model

$$Y_{ijkd} = \mu + \delta_i + \pi_j + s_{jk} + e_{ijkd} \quad (2.2.9)$$

where $d = 1, \dots, n_{ijk}$, $k = 1, \dots, m_j$, $i = 1, 2, 3$ and $j = 1, \dots, 5$ where n_{ijk} is the number of lambs whose dams belong to the i^{th} age category when the population line is j and the sire is k and m_j is the total number of lambs whose sires are from population line j . Here the age effects $(\delta_1, \delta_2, \delta_3)$ and line effects $(\pi_1, \dots, \pi_5)$ are considered as fixed effects and the sire (within line) effects s_{jk} are iid $N(0, (r\lambda)^{-1})$ and independent of error variables e_{ijkd} which are iid $N(0, r^{-1})$. To make the design matrix associated with fixed effects full rank we can take $\delta_3 = 0 = \pi_5$ which is the usual formulation needed for GLM Procedures in SAS.

Let $\mu_{ij} = E(Y_{ijkd}) = \mu + \delta_i + \pi_j$ and let there be $n_{i..}$ observations corresponding to the i^{th} age category. We will be interested in predicting

$$\begin{aligned} w_{jk} &= \frac{1}{n_T} \sum_{i=1}^3 \mu_{ij} n_{i..} + s_{jk} \\ &= \mu + \frac{1}{n_T} \sum_{i=1}^3 n_{i..} \delta_i + \pi_j + s_{jk} \end{aligned} \quad (2.2.10)$$

where $n_T = \sum_{i=1}^3 n_{i..}$. The value of w_{jk} can be interpreted as the average birth weight of an infinite number of male lambs that are offspring of the k^{th} sire in the j^{th} line. Clearly, this is a special case of the general problem we are interested in with appropriately defined b , v , X , Z , S and T .

2.3 Hierarchical Bayes Analysis

In this section, for the finite population sampling we provide the predictive distribution of $Y^{(2)}$ given $Y^{(1)} = y^{(1)}$ and for the infinite population set up we provide the posterior distribution of the vector of effects $(b^T, y^T)^T$ given $Y = y$.

We will use the following notations to label certain distributions to be used in this section. A random variable Z is said to have a gamma(α, β) distribution if it has pdf

$$f(z) = \left[\exp(-\alpha z) z^{\beta-1} \alpha^\beta / \Gamma(\beta) \right] I_{[z>0]}. \quad (2.3.1)$$

A random vector $T = (T_1, \dots, T_p)^T$ is said to have a multivariate t-distribution with location parameter $\underline{\mu}$, scale parameter Φ , a p.d. p x p matrix and degrees of freedom (d.f.) ν if it has pdf

$$g(\underline{t}) \propto |\Phi|^{-\frac{1}{2}} \left[\nu + (\underline{t} - \underline{\mu})^T \Phi^{-1} (\underline{t} - \underline{\mu}) \right]^{-\frac{1}{2}(\nu+p)}, \quad (2.3.2)$$

(see Zellner, 1971, p. 383, or Press, 1972, p. 136). Here $|E|$ denotes the determinant of a square matrix E . Assume $\nu > 2$. Then $E(T) = \underline{\mu}$, $V(T) = (\nu/(\nu-2))\Phi$.

We assume conditions (A) and (B) given at the beginning of Section 2.2. In stage (C) of the model, it is assumed that

(C1) $B, R, \Lambda_1 R, \dots, \Lambda_t R$ are independently distributed with $B \sim \text{uniform}(R^P)$, $R \sim \text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$, $a_0 \geq 0$, $g_0 \geq 0$, $\Lambda_i R \sim \text{gamma}(\frac{1}{2}a_i, \frac{1}{2}g_i)$ with $a_i > 0$, $g_i \geq 0$ ($i = 1, \dots, t$).

Allowing a_0 and some of g_i to be zero, some improper gamma distributions are included as a possibility in our prior.

Before stating the predictive distribution of $Y^{(2)}$ given $Y^{(1)} = \underline{y}^{(1)}$, we need to introduce a few matrix notations. We write $\underline{\Sigma} \equiv \underline{\Sigma}(\lambda) = \Psi + \underline{Z}D(\lambda)\underline{Z}^T$, and partition $\underline{\Sigma}$ into $\underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{bmatrix}$. Also, let

$$\underline{\Sigma}_{22.1} = \underline{\Sigma}_{22} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{12}; \quad (2.3.3)$$

$$\underline{K} = \underline{\Sigma}_{11}^{-1} - \underline{\Sigma}_{11}^{-1}\underline{X}^{(1)}(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})^{-1}\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}; \quad (2.3.4)$$

$$\underline{M} = \underline{\Sigma}_{21}\underline{K} + \underline{X}^{(2)}(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})^{-1}\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}; \quad (2.3.5)$$

$$\begin{aligned} \underline{G} = \underline{\Sigma}_{22.1} + (\underline{X}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})^{-1} \\ \times (\underline{X}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})^T. \end{aligned} \quad (2.3.6)$$

Now the predictive distribution of $Y^{(2)}$ given $Y^{(1)} = \underline{y}^{(1)}$ is given in the following theorem in two steps. A proof of this theorem is deferred to Appendix A.

Theorem 2.3.1. Consider the model given in (2.2.2) and

(C1). Assume that $n_T + \sum_{i=0}^t g_i - p > 2$. Then, conditional

on $\underline{\Lambda} = \underline{\lambda}$ and $Y^{(1)} = \underline{y}^{(1)}$, $Y^{(2)}$ has multivariate t-

distribution with d.f. $n_T + \sum_{i=0}^t g_i - p$, location parameter

$\underline{M}_Y^{(1)}$, and scale parameter $\left(n_T + \sum_{i=0}^t g_i - p\right)^{-1} \times$

$\left[a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K}_Y^{(1)} \underline{y}^{(1)}\right] \underline{G}$. Also, the conditional

distribution of $\underline{\Lambda}$ given $Y^{(1)} = \underline{y}^{(1)}$ has pdf

$$f(\underline{\lambda} | \underline{y}^{(1)}) \propto |\underline{\Sigma}_{11}|^{-\frac{1}{2}} \left| \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right|^{-\frac{1}{2}} \left[\prod_{i=1}^t \lambda_i^{\frac{1}{2} g_i - 1} \right] \\ \times \left[a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K}_Y^{(1)} \underline{y}^{(1)} \right]^{-\frac{1}{2} (n_T + \sum_{i=0}^t g_i - p)}. \quad (2.3.7)$$

Using the moments of a multivariate t-distribution and the iterated formulas for conditional expectations and variances it follows from the above theorem that

$$E(Y^{(2)} | \underline{y}^{(1)}) = E(\underline{M} | \underline{y}^{(1)}) \underline{y}^{(1)}; \quad (2.3.8)$$

$$V(Y^{(2)} | \underline{y}^{(1)}) \\ = V\left(E(Y^{(2)} | \underline{\Lambda}, \underline{y}^{(1)}) | \underline{y}^{(1)}\right) + E\left(V(Y^{(2)} | \underline{\Lambda}, \underline{y}^{(1)}) | \underline{y}^{(1)}\right)$$

$$\begin{aligned}
&= v(\underline{M}\underline{y}^{(1)}|\underline{y}^{(1)}) + \left(n_T + \sum_{i=0}^t g_i - p - 2\right)^{-1} \\
&\times E\left[\left\{a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K}\underline{y}^{(1)}\right\} \underline{G} \middle| \underline{y}^{(1)}\right]. \quad (2.3.9)
\end{aligned}$$

Using (2.3.8) and (2.3.9), it is possible to find the posterior mean and variance of $\underline{\xi} \equiv \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) = \underline{A}\underline{Y}^{(1)} + \underline{C}\underline{Y}^{(2)}$ where $\underline{A}$ and $\underline{C}$ are known matrices. The Bayes estimate of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ under any quadratic loss is its posterior mean, and is given by

$$\underline{e}_{BF}(\underline{y}^{(1)}) = \left[\underline{A} + \underline{C}E(\underline{M}|\underline{y}^{(1)})\right]\underline{y}^{(1)}, \quad (2.3.10)$$

using (2.3.8). Similarly, using (2.3.9), one may obtain

$$v\left[\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) \middle| \underline{y}^{(1)}\right] = \underline{C}v(\underline{Y}^{(2)}|\underline{y}^{(1)})\underline{C}^T. \quad (2.3.11)$$

Note that when $\underline{A} = \bigoplus_{i=1}^m \underline{1}_{n_i}^T$ and $\underline{C} = \bigoplus_{i=1}^m \underline{1}_{N_i-n_i}^T$, $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ reduces to the vector of finite population totals for the m small areas, whereas for the choice $\underline{A} = \bigoplus_{i=1}^m (\underline{1}_{n_i}^T/N_i)$ and $\underline{C} = \bigoplus_{i=1}^m (\underline{1}_{N_i-n_i}^T/N_i)$, $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ reduces to the vector of finite population means for the m small areas.

Now we will get back to the infinite population set up to provide the posterior distribution of $\underline{W} = (\underline{b}^T, \underline{y}^T)^T$ given $\underline{Y} = \underline{y}$ which will be used to find the posterior mean and variance of $\underline{\zeta}(\underline{b}, \underline{y})$, a vector of the linear combination of $\underline{W}$. The posterior distribution is given in the following

theorem. A proof of this theorem will be omitted because of its similarity to the proof of Theorem 2.3.1. We will consider the model given by (A) and (B) of Section 2.2 and (C1). Recall from the middle of Section 2.2 that we have redefined the dimensions of $\underline{Y}$, $\underline{X}$, $\underline{Z}$ and $\underline{e}$ appearing there by $\underline{Y}(n_T \times 1)$, $\underline{X}(n_T \times p)$, $\underline{Z}(n_T \times q)$ and $\underline{e}(n_T \times 1)$. Also we have assumed $\text{rank}(\underline{X}) = p$. Now we will state the theorem.

Theorem 2.3.2. Consider the model stated above and assume that $n_T + \sum_{i=0}^t g_i - p > 2$. Then, conditional on $\underline{\Lambda} = \underline{\lambda}$ and $\underline{Y} = \underline{y}$, $\underline{W}$ has multivariate t-distribution with d.f.

$n_T + \sum_{i=0}^t g_i - p$, location parameter $\underline{P}\underline{y}$, scale parameter $\left(n_T + \sum_{i=0}^t g_i - p\right)^{-1} \left[a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^T \underline{Q} \underline{y} \right] \underline{H}$, where

$$\underline{Q} = \underline{\Sigma}^{-1} - \underline{\Sigma}^{-1} \underline{X} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1}; \quad (2.3.12)$$

$$\underline{P}^T = \left[\underline{\Sigma}^{-1} \underline{X} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \quad \underline{Q} \underline{Z} \underline{D} \right]; \quad (2.3.13)$$

$$\underline{H} = \begin{bmatrix} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} & -(\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1} \underline{Z} \underline{D} \\ -\underline{D} \underline{Z}^T \underline{\Sigma}^{-1} \underline{X} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} & \underline{D} - \underline{D} \underline{Z}^T \underline{Q} \underline{Z} \underline{D} \end{bmatrix}. \quad (2.3.14)$$

Also, the conditional distribution of $\underline{\Lambda}$ given $\underline{Y} = \underline{y}$ has pdf

$$f(\underline{\lambda}|\underline{y}) \propto |\underline{\Sigma}|^{-\frac{1}{2}} |\underline{X}^T \underline{\Sigma}^{-1} \underline{X}|^{-\frac{1}{2}} \left[\prod_{i=1}^t \lambda_i^{\frac{1}{2}} g_i^{-1} \right] \\ \times \left[a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^T \underline{Q} \underline{y} \right]^{-\frac{1}{2} \left(n_T + \sum_{i=0}^t g_i^{-p} \right)}. \quad (2.3.15)$$

Again using the moments of multivariate t-distribution, and the iterated formulas for expectation and variance, the above theorem can be used to find the computational formulas for $E(\underline{W}|\underline{y})$ and $V(\underline{W}|\underline{y})$ as in (2.3.8) and (2.3.9). Similarly, one can find $E(\zeta(\underline{b}, \underline{y})|\underline{y}) = e_{BI}(\underline{y})$ (say) and $V(\zeta(\underline{b}, \underline{y})|\underline{y})$ as in (2.3.10) and (2.3.11) where $\zeta(\underline{b}, \underline{y}) = \underline{S}\underline{b} + \underline{T}\underline{y}$ for known matrices $\underline{S}(uxp)$ and $\underline{T}(uxq)$.

Applications of these two theorems will be considered in Section 2.4 to some actual data sets. There we will carry out an HB analysis of the data sets which appeared in Battese et al. (1988) and Harville (in press).

Before we conclude this section, we will make a final observation. A comparison of (2.3.4) and (2.3.7) with (2.3.12) and (2.3.15) reveals that if we replace $\underline{y}$ by $\underline{y}^{(1)}$, $\underline{X}$ by $\underline{X}^{(1)}$ and $\underline{\Sigma}$ by $\underline{\Sigma}_{11}$ in $f(\underline{\lambda}|\underline{y})$ in (2.3.15) we obtain $f(\underline{\lambda}|\underline{y}^{(1)})$ as given in (2.3.7). This observation will be referred to in Section 2.4.

2.4 Applications of Hierarchical Bayes Analysis

This section concerns the analysis of two real data sets using the HB procedures suggested in Section 2.3. The first data set is related to the prediction of corn and soybeans for 12 counties in north-central Iowa based on 1978 June Enumerative Survey as well as LANDSAT satellite data. It appeared in Battese, Harter and Fuller (BHF) who conducted a variance components analysis for this problem. The second data set originally appeared in Harville and Fenech (1985) and reappeared in Harville (in press) where he conducted a variance components as well as an HB analysis to predict w_{jk} , given in (2.2.10), the average weight of an infinite number of single-birth male lambs that are offspring of the k^{th} sire in the j^{th} population line.

We will first consider the BHF data set. To start with, we briefly give a background of this problem. The USDA Statistical Reporting Service field staff determined the area of corn and soybeans in 37 sample segments (each segment about 250 hectares) of 12 counties in north-central Iowa by interviewing farm operators. Based on LANDSAT readings obtained during August and September 1978, USDA procedures were used to classify the crop cover for all pixels (a term for "picture element" about .45 hectares) in the 12 counties. The number of segments in each county,

the number of hectares of corn and soybeans (as reported in the June Enumerative Survey), the number of pixels classified as corn and soybeans for each sample segment, and the county mean number of pixels classified as corn and soybeans (the total number of pixels classified as that crop divided by the number of segments in that county) are reported in Table 1 of BHF. For ready reference, it is reproduced in Table 2.1. In order to make our results comparable to that of BHF, the second segment in Hardin county was ignored.

The model considered by BHF is

$$Y_{ij} = b_0 + b_1 x_{1ij} + b_2 x_{2ij} + v_i + e_{ij}, \quad (2.4.1)$$

where i is a subscript for the county, and j is a subscript for a segment within the given county ($j = 1, \dots, N_i$, the number of segments in the i^{th} county, $i = 1, \dots, 12$). Here x_{1ij} is the number of pixels of corn and x_{2ij} is the number of pixels of soybeans for the j^{th} segment in the i^{th} county. They assumed (in our notations) $E(v_i) = E(e_{ij}) = 0$, $V(v_i) = (\lambda r)^{-1}$, $V(e_{ij}) = r^{-1}$, $\text{Cov}(v_i, e_{ij}) = 0$, $\text{Cov}(v_i, v_{i'}) = 0$ ($i \neq i'$), $\text{Cov}(e_{ij}, e_{i'j'}) = 0$ if $(i, j) \neq (i', j')$. As in BHF, we are interested in predicting $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$, the finite population mean for the i^{th} county both for corn and soybeans. Using (2.4.1),

Table 2.1 Survey and Satellite Data for Corn and Soybeans
in 12 Iowa Counties

County	Sample	County	No. of		No. of pixels		Mean no. of	
			Segments	Reported hectares	in sample segments	in sample segments	pixels per segment*	pixels per segment*
			Corn	Soybean	Corn	Soybean	Corn	Soybean
Cerro Gordo	1	545	165.76	8.09	374	55	295.29	189.70
Hamilton	1	566	96.32	106.03	209	218	300.40	196.65
Worth	1	394	76.08	103.60	253	250	289.60	205.28
Humboldt	2	424	185.35	6.47	432	96	290.74	220.22
			116.43	63.82	367	178		
Franklin	3	564	162.08	43.50	361	137	318.21	188.06
			152.04	71.43	288	206		
			161.75	42.49	369	165		
Pocahontas	3	570	92.88	105.26	206	218	257.17	247.13
			149.94	76.49	316	221		
			64.75	174.34	145	338		
Winnebago	3	402	127.07	95.67	355	128	291.77	185.37
			133.55	76.57	295	147		
			77.70	93.48	223	204		
Wright	3	567	206.39	37.84	459	77	301.26	221.36
			108.33	131.12	290	217		
			118.17	124.44	307	258		
Webster	4	687	99.96	144.15	252	303	262.17	247.09
			140.43	103.60	293	221		
			98.95	88.59	206	222		
			131.04	115.58	302	274		
Hancock	5	569	114.12	99.15	313	190	314.28	198.66
			100.60	124.56	246	270		
			127.88	110.88	353	172		
			116.90	109.14	271	228		
			87.41	143.66	237	297		
Kossuth	5	965	93.48	91.05	221	167	298.65	204.61
			121.00	132.33	369	191		
			109.91	143.14	343	249		
			122.66	104.13	342	182		
			104.21	118.57	294	179		
Hardin	6	556	88.59	102.59	220	262	325.99	177.05
			88.59	29.46	340	87		
			165.35	69.28	355	160		
			104.00	99.15	261	221		
			88.63	143.66	187	345		
			153.70	94.49	350	190		

*The mean number of pixels of a given crop per segment in a county is the total number of pixels classified as that crop, divided by the number of segments in that county.

$\bar{Y}_i$ can be written as $\bar{Y}_i = \mu_i + \bar{e}_i$ where $\bar{e}_i = N_i^{-1} \sum_{j=1}^{N_i} e_{ij}$, $\mu_i = b_0 + b_1 \bar{x}_{1i}(p) + b_2 \bar{x}_{2i}(p) + v_i$, $\bar{x}_{1i}(p) = N_i^{-1} \sum_{j=1}^{N_i} x_{1ij}$ and $\bar{x}_{2i}(p) = N_i^{-1} \sum_{j=1}^{N_i} x_{2ij}$. Under the assumptions of model (2.4.1), μ_i can be interpreted as the conditional mean of the hectares of corn (or soybeans) per segment, given the realized county effect v_i and the values of the satellite data. Clearly, the mean $\bar{Y}_i$ is not equivalent to μ_i , because the average of the e_{ij} over the finite population of segments in county i is not identically 0. However, either if N_i are large or if the sampling rates n_i/N_i ($i = 1, \dots, m$) are small, then the predictor of μ_i is an appropriate predictor of $\bar{Y}_i$. In this example, either of the conditions appears to be true. For predicting $\bar{Y}_i$, first assuming λ and r known, BHF obtained BLUPs of μ_i ($i = 1, \dots, 12$). Then, using Henderson's Method III, they obtained estimates of the variance components, and the final predictors involved the estimated variance components. Henderson's method being an ANOVA method could lead to negative estimates of λ^{-1} . If this were the case, BHF set it equal to zero. This phenomenon is likely to happen, particularly when the number of small areas or strata is small.

We use instead Theorem 2.3.1 and Theorem 2.3.2 respectively to find the first two posterior moments of $\bar{Y}_i$

and μ_i . In the special case of nested error regression model, we will now develop the expressions for the posterior distribution given by (2.3.7) and the expressions for the posterior means and variances of $\bar{Y}_i$ and μ_i . Here, we have $t = 1$, $\lambda_1 = \lambda$, $D(\lambda) = \lambda^{-1} I_m$, $\Psi = I_{N_T}$. Then $\Sigma_{11} = \bigoplus_{i=1}^m (I_{n_i} + \lambda^{-1} J_{n_i})$ so that $|\Sigma_{11}| = \prod_{i=1}^m \{(\lambda + n_i)/\lambda\}$. Also, writing $\bar{x}_i(s) = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ ($i = 1, \dots, m$) where $x_{ij} = (1 \ x_{1ij} \ x_{2ij})^T$, one gets

$$\begin{aligned} & \underline{X}^{(1)T} \Sigma_{11}^{-1} \underline{X}^{(1)} \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} x_{ij}^T - \sum_{i=1}^m n_i^2 (n_i + \lambda)^{-1} \bar{x}_i(s) \bar{x}_i^T(s) \\ &= H(\lambda) \text{ (say)}. \end{aligned} \quad (2.4.2)$$

Next writing $\bar{y}_i(s) = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$, one gets

$$\begin{aligned} & \underline{Y}^{(1)T} K_Y \underline{Y}^{(1)} \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i(s))^2 + \lambda \sum_{i=1}^m n_i (n_i + \lambda)^{-1} \bar{y}_i^2(s) \\ &\quad - \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_i(s)) \right\}^T \\ &\quad \times H^{-1}(\lambda) \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_i(s)) \right\} \\ &= Q_0(\lambda) \text{ (say)}. \end{aligned} \quad (2.4.3)$$

The conditional pdf $f(\lambda | \underline{y}^{(1)})$ given in (2.3.7) simplifies to

$$f(\lambda | \underline{y}^{(1)}) \propto \lambda^{\frac{1}{2}(m+g_1)-1} \prod_{i=1}^m (\lambda + n_i)^{-\frac{1}{2}} |\underline{H}(\lambda)|^{-\frac{1}{2}} \\ (a_0 + a_1 \lambda + Q_0(\lambda))^{-\frac{1}{2}(n_T+g_0+g_1-p)} \quad (2.4.4)$$

Next writing $f_i = (N_i - n_i)/N_i$, $\bar{x}_i^* = (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} x_{ij}$, the posterior means, variances and covariances of the finite population means are given by

$$E \left[N_i^{-1} \sum_{j=1}^{N_i} Y_{ij} \middle| \underline{y}^{(1)} \right] \\ = (1 - f_i) \bar{y}_{i(s)} + f_i E \left[n_i (n_i + \lambda)^{-1} \bar{y}_{i(s)} \middle| \underline{y}^{(1)} \right] \\ + f_i E \left[\left\{ \bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_{i(s)} \right\}^T \underline{H}^{-1}(\lambda) \right. \\ \left. \times \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_{i(s)}) \right\} \middle| \underline{y}^{(1)} \right] \\ = e_{HB} \quad (\text{say}); \quad (2.4.5)$$

$$V \left[N_i^{-1} \sum_{j=1}^{N_i} Y_{ij} \middle| \underline{y}^{(1)} \right] \\ = f_i^2 V \left[n_i (n_i + \lambda)^{-1} \bar{y}_{i(s)} + \left\{ \bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_{i(s)} \right\}^T \right. \\ \left. \times \underline{H}^{-1}(\lambda) \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_{i(s)}) \right\} \middle| \underline{y}^{(1)} \right] \\ + (n_T + g_0 + g_1 - p - 2)^{-1} f_i E \left[\left\{ a_0 + a_1 \lambda + Q_0(\lambda) \right\} \right]$$

$$\begin{aligned}
& \times \left\{ \left(N_i + \lambda \right) \left(N_i (n_i + \lambda) \right)^{-1} + f_i \left(\bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right)^T \right. \\
& \times \left. H^{-1}(\lambda) \left(\bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right) \right\} \Big| \underline{y}^{(1)} \Big] \\
& = V_1 + V_2 \quad (\text{say}) \\
& = s_{HB}^2 \quad (\text{say}); \tag{2.4.6}
\end{aligned}$$

and for $i \neq k$,

$$\begin{aligned}
& \text{Cov} \left[N_i^{-1} \Sigma_{j=1}^{N_i} Y_{ij}, N_k^{-1} \Sigma_{j=1}^{N_k} Y_{kj} \Big| \underline{y}^{(1)} \right] \\
& = f_i f_k \text{Cov} \left[n_i (n_i + \lambda)^{-1} \bar{y}_i(s) + \left\{ \bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right\}^T \right. \\
& \times H^{-1}(\lambda) \left\{ \Sigma_{i=1}^m \Sigma_{j=1}^{n_i} \bar{x}_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_i(s)) \right\}, \\
& n_k (n_k + \lambda)^{-1} \bar{y}_k(s) + \left\{ \bar{x}_k^* - n_k (n_k + \lambda)^{-1} \bar{x}_k(s) \right\}^T H^{-1}(\lambda) \\
& \times \left. \left\{ \Sigma_{k=1}^m \Sigma_{j=1}^{n_k} \bar{x}_{kj} (y_{kj} - n_k (n_k + \lambda)^{-1} \bar{y}_k(s)) \right\} \Big| \underline{y}^{(1)} \right] \\
& + (n_T + g_0 + g_1 - p - 2)^{-1} f_i f_k E \left[\left(a_0 + a_1 \lambda + Q_0(\lambda) \right) \right. \\
& \times \left. \left(\bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right)^T H^{-1}(\lambda) \left(\bar{x}_k^* - n_k (n_k + \lambda)^{-1} \bar{x}_k(s) \right) \Big| \underline{y}^{(1)} \right]. \tag{2.4.7}
\end{aligned}$$

Although we have provided the expressions for the

covariances which may be necessary for providing simultaneous confidence set for the finite population mean vector, we will not use it here.

Before we find the posterior means and variances of μ_i in the infinite population set up, we give a general discussion comparing HB predictors with EB predictors. Writing

$$\begin{aligned}
 & E \left[N_i^{-1} \sum_{j=1}^{N_i} Y_{ij} \middle| \lambda, \underline{y}^{(1)} \right] \\
 &= (1 - f_i) \bar{y}_i(s) + f_i n_i (n_i + \lambda)^{-1} \bar{y}_i(s) \\
 &\quad + f_i \left\{ \bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right\}^T H^{-1}(\lambda) \\
 &\quad \times \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} (y_{ij} - n_i (n_i + \lambda)^{-1} \bar{y}_i(s)) \right\} \\
 &= g_1(\lambda) \quad (\text{say}), \tag{2.4.8}
 \end{aligned}$$

and

$$\begin{aligned}
 & V \left[N_i^{-1} \sum_{j=1}^{N_i} Y_{ij} \middle| \lambda, \underline{y}^{(1)} \right] \\
 &= (n_T + g_0 + g_1 - p - 2)^{-1} f_i \{ a_0 + a_1 \lambda + Q_0(\lambda) \} \\
 &\quad \times \left\{ (N_i + \lambda) (N_i (n_i + \lambda))^{-1} + f_i \left(\bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right)^T \right. \\
 &\quad \times H^{-1}(\lambda) \left(\bar{x}_i^* - n_i (n_i + \lambda)^{-1} \bar{x}_i(s) \right) \left. \right\}
 \end{aligned}$$

$$= g_2(\lambda) \quad (\text{say}), \quad (2.4.9)$$

we have from (2.4.5), (2.4.6), (2.4.8) and (2.4.9) that

$$e_{HB} = E[g_1(\lambda)|y^{(1)}], \quad (2.4.10)$$

$$V_1 = V[g_1(\lambda)|y^{(1)}], \quad (2.4.11)$$

and

$$V_2 = E[g_2(\lambda)|y^{(1)}]. \quad (2.4.12)$$

In EB analysis, to obtain the EB predictor, we usually replace $E[g_1(\lambda)|y^{(1)}]$ by $g_1(\hat{\lambda}) = e_{EB}$ (say) where $\hat{\lambda}$ is some estimate of λ , which can be ML, REML or ANOVA estimate and we report a naive measure of posterior variance by $g_2(\hat{\lambda}) = s_{EB}^2$ (say). Usually, the point estimates e_{HB} and e_{EB} are not too far apart. But to measure the posterior variance s_{HB}^2 by s_{EB}^2 , we may underestimate the actual measure because of the failure to account for the estimation of λ . We may grossly underestimate the actual measure if $g_1(\lambda)$ varies too much within the body of the posterior distribution of λ as in this case V_1 will be significantly large. We will see in this example that for some of the counties V_1 contributes a significant portion of the total variance. We will also find that the percent contribution of V_1 to

s_{HB}^2 usually increases with the relative difference between e_{HB} and e_{EB} .

Now to develop the expressions for the posterior means and variances for μ_i we use Theorem 2.3.2. Note that by an observation made at the end of Section 2.3, $f(\lambda|\underline{y})$, the posterior distribution of λ given $\underline{y}$, is given by (2.4.4). After considerable simplifications in this particular case, we obtain

$$\begin{aligned} E[\mu_i|\underline{y}] &= E\left[n_i(n_i + \lambda)^{-1}\bar{y}_i(s)|\underline{y}\right] \\ &+ E\left[\left\{\bar{x}_i(p) - n_i(n_i + \lambda)^{-1}\bar{x}_i(s)\right\}^T H^{-1}(\lambda)\right. \\ &\times \left.\left\{\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}(y_{ij} - n_i(n_i + \lambda)^{-1}\bar{y}_i(s))\right\}|\underline{y}\right] \\ &= e_{HB}^* \quad (\text{say}); \end{aligned} \quad (2.4.13)$$

$$\begin{aligned} V[\mu_i|\underline{y}] &= V\left[n_i(n_i + \lambda)^{-1}\bar{y}_i(s) + \left\{\bar{x}_i(p) - n_i(n_i + \lambda)^{-1}\bar{x}_i(s)\right\}^T\right. \\ &\times \left.H^{-1}(\lambda)\left\{\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}(y_{ij} - n_i(n_i + \lambda)^{-1}\bar{y}_i(s))\right\}|\underline{y}\right] \\ &+ (n_T + g_0 + g_1 - p - 2)^{-1}E\left[\left\{a_0 + a_1\lambda + Q_0(\lambda)\right\}\right. \\ &\times \left.\left\{(n_i + \lambda)^{-1} + \left(\bar{x}_i(p) - n_i(n_i + \lambda)^{-1}\bar{x}_i(s)\right)^T\right\}\right. \\ &\times \left.H^{-1}(\lambda)\left(\bar{x}_i(p) - n_i(n_i + \lambda)^{-1}\bar{x}_i(s)\right)\right]|\underline{y} \end{aligned}$$

$$= s_{HB}^{*2} \quad (\text{say}); \quad (2.4.14)$$

where $\bar{x}_i(p) = \left(1, \bar{x}_{1i}(p), \bar{x}_{2i}(p) \right)^T$, $H(\lambda)$ as given in (2.4.2) and $Q_0(\lambda)$ as given in (2.4.3). Note that since $f_i \rightarrow 1$ and $\bar{x}_i(p) - \bar{x}_i^* \rightarrow 0$ as $N_i \rightarrow \infty$, it can be seen informally that the rhs of (2.4.5) and (2.4.6) approach in limit (2.4.13) and (2.4.14), respectively.

We will now get back to the actual data analysis of the BHF data set given in Table 2.1. We use formulas (2.4.5), (2.4.6), (2.4.8), (2.4.9), (2.4.13) and (2.4.14) to obtain HB and EB posterior means and variances of the population means for the 12 counties. Our HB approach eliminates the possibility of obtaining zero estimates of the variance components. A number of different priors for R and RA were tried; both informative and noninformative. The results for the posterior means were quite similar whereas the posterior variances varied approximately by as much as 10%. For illustration purpose, we have decided to report our analysis for the prior with $a_0 = 0.005$, $g_0 = 0$, $a_1 = 0.005$ and $g_1 = 0$. But since the choice of $a_1 = 0$ gives improper posterior distribution of λ , we took a_1 a small positive number. Table 2.2 provides the HB predictors e_{HB} , the EB predictors e_{EB} , the BHF predictors

Table 2.2 The Predicted Hectares of Corn and Associated Standard Errors

	$a_0 = .005$	$g_0 = 0$	$a_1 = .005$	$g_1 = 0$		
County	e_{HB}	e_{EB}	e_{BHF}	s_{HB}	s_{EB}	s_{BHF}
Cerro Gordo	122.1	122.2	122.2	9.3	9.4	10.3
Franklin	143.6	144.2	145.3	6.9	6.4	6.7
Hamilton	126.2	126.2	126.5	9.2	9.3	10.1
Hancock	124.6	124.4	124.2	5.3	5.3	5.5
Hardin	142.6	143.0	143.5	5.8	5.6	5.8
Humboldt	108.9	108.5	107.7	8.2	7.9	8.4
Kossuth	107.7	106.9	106.1	5.8	5.2	5.4
Pocahontas	111.8	112.1	112.9	6.6	6.4	6.8
Webster	114.9	115.3	116.0	5.9	5.7	6.0
Winnebago	113.3	112.8	112.1	6.6	6.4	6.8
Worth	107.1	106.8	105.6	9.9	9.1	10.0
Wright	122.0	122.0	122.1	6.4	6.5	6.9

e_{BHF} and the respective associated standard errors s_{HB} , s_{EB} and s_{BHF} for the corn data. Table 2.3 provides the values of e_{HB} , e_{EB} , e_{BHF} and e_{HB}^* for the soybeans data for the same choice of prior hyperparameters, whereas Table 2.4 provides their respective standard errors along with the components V_1 and V_2 of s_{HB}^2 . Values of e_{BHF} and s_{BHF} presented in Tables 2.2 - 2.4 are computed using FORTRAN from the formulas given in the BHF paper and are slightly different from the values reported in Battese et al. (1988). From Tables 2.2 and 2.3, for predicting corn and soybeans, one can see that e_{HB} , e_{HB}^* , e_{EB} and e_{BHF} are quite close to each other. From Tables 2.2 and 2.4, s_{EB} and s_{HB} appear to be smaller than s_{BHF} . But since s_{EB} is naive EB posterior s.d., it is probably an underestimate of the true measure. From Tables 2.3 and 2.4 we find hardly any difference either between e_{HB} and e_{HB}^* or between their standard errors s_{HB} and s_{HB}^* . This is what we anticipated for this data. To draw a clear comparison between HB and EB procedures, we added one extra column at the end of Tables 2.3 and 2.4. The last column of Table 2.3 measures the percent relative difference $100 \times |e_{\text{HB}} - e_{\text{EB}}|/e_{\text{HB}} \%$ between EB and HB predicted values whereas the last column of Table 2.4 measures the percent contribution $(100 \times V_1/(V_1 + V_2)\%)$ of V_1 towards the total posterior variance s_{HB}^2 . Comparison of these two columns indicates that the

Table 2.3 The Predicted Hectares of Soybeans Obtained by
Using Different Procedures

	$a_0 = .005$	$g_0 = 0$	$a_1 = .005$	$g_1 = 0$	
County	e_{HB}	e_{HB}^*	e_{EB}	e_{BHF}	$ 1 - e_{EB}/e_{HB} \times 100\%$
Cerro Gordo	78.8	78.8	78.2	77.5	0.78
Franklin	67.1	67.1	65.9	64.8	1.80
Hamilton	94.4	94.4	94.6	96.0	0.21
Hancock	100.4	100.4	100.8	101.1	0.40
Hardin	75.4	75.4	75.1	74.9	0.39
Humboldt	81.9	82.0	80.6	79.2	1.71
Kossuth	118.2	118.2	119.2	120.2	0.84
Pocahontas	113.9	113.9	113.7	113.8	0.18
Webster	110.0	110.0	109.7	109.6	0.37
Winnebago	97.3	97.3	98.0	98.7	0.72
Worth	87.8	87.8	87.2	86.6	0.68
Wright	111.9	111.9	112.4	112.9	0.45

Table 2.4 The Standard Errors Associated with Different Predictors of Hectares of Soybeans

	$a_0 = .005$		$g_0 = 0$		$a_1 = .005$		$g_1 = 0$	
County	s_{HB}	s_{HB}^*	s_{EB}	s_{BHF}	V_1	V_2	$V_1/(V_1+V_2) \times 100\%$	
Cerro Gordo	11.7	11.7	11.6	12.7	7.67	128.59	5.1	
Franklin	8.2	8.2	7.5	7.8	11.94	54.92	18.0	
Hamilton	11.2	11.2	11.4	12.4	1.97	123.61	1.6	
Hancock	6.2	6.3	6.1	6.3	1.35	37.59	3.4	
Hardin	6.5	6.5	6.5	6.6	0.37	41.84	0.9	
Humboldt	10.4	10.4	9.9	10.0	22.62	85.40	20.9	
Kossuth	6.6	6.7	6.0	6.2	7.99	36.23	18.1	
Pocahontas	7.5	7.5	7.5	7.9	0.06	55.98	0.1	
Webster	6.6	6.7	6.6	6.8	0.64	43.51	1.5	
Winnebago	7.7	7.8	7.5	7.9	4.11	55.70	6.9	
Worth	11.1	11.1	11.1	12.1	4.06	118.17	3.3	
Wright	7.7	7.7	7.6	8.0	1.62	57.48	2.7	

contribution of V_1 usually increases with the relative difference. In particular, for counties Franklin, Humboldt and Kossuth these relative differences are as high as 1.80%, 1.71% and .84% and the corresponding contributions of V_1 are as nonnegligible as 18.0%, 20.9% and 18.1%. This made s_{EB} much smaller than s_{HB} for these counties. So if one uses a naive EB or estimated BLUP approach, he will tend to underestimate the mean squared error (MSE) of prediction. One should note that though BHF used estimated BLUP, they tried to account for the uncertainty involved in the estimation of λ in their approximations of MSE. Similar approximations of MSE of prediction have been suggested by Kackar and Harville (1984), Prasad and Rao (1990) and Lahiri and Rao (1990).

Now we will consider the lamb-weight data set of Harville (in press). The background of the data set is given in the example presented at the end of Section 2.2. We will use a model similar to the one given in (2.2.9) to analyze the data set. There we assumed, following Harville (in press), the population line effects as fixed. For the purpose of illustration with three variance components, we will assume the population effects as random. This would have been appropriate if these five lines were randomly selected from a large number of populations lines. Also to make the design matrix associated with the fixed effects

(age of dam) of full column rank, we will write $\mu + \delta_i$ as μ_i . Now we have the following mixed linear model

$$Y_{ijkd} = \mu_i + \pi_j + s_{jk} + e_{ijkd} \quad (2.4.15)$$

$d = 1, \dots, n_{ijk}$, $k = 1, \dots, m_j$, $i = 1, 2, 3$ and $j = 1, \dots, 5$ where n_{ijk} and m_j are the same as in (2.2.9). We assume π_j are iid $N(0, (r\lambda_1)^{-1})$, s_{jk} are iid $N(0, (r\lambda_2)^{-1})$ and e_{ijkd} are iid $N(0, r^{-1})$. Moreover π_j , s_{jk} and e_{ijkd} are assumed to be mutually independent. We want to predict w_{jk} given in (2.2.10). Using (2.4.15), we will rewrite w_{jk} as

$$w_{jk} = \frac{1}{n_T} \sum_{i=1}^3 n_{i..} \mu_i + \pi_j + s_{jk} \quad (2.4.16)$$

where $n_{i..}$ and n_T are given in (2.2.10).

We will carry out a noninformative Bayesian analysis using a uniform(R^3) prior for $\underline{\mu} = (\mu_1, \mu_2, \mu_3)^T$ and independent $\text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$, $\text{gamma}(\frac{1}{2}a_1, \frac{1}{2}g_1)$ and $\text{gamma}(\frac{1}{2}a_2, \frac{1}{2}g_2)$ priors for R , $R\lambda_1$ and $R\lambda_2$ respectively. Using Theorem 2.3.2, w_{jk} being a linear combination of fixed and random effects, we can find its posterior mean $E(w_{jk}|\underline{y}) = e_{HB}^*$ (say) and the posterior variance $V(w_{jk}|\underline{y}) = s_{HB}^{*2}$ (say). Using this theorem and iterated formulas for mean and variance, we can derive expressions for e_{HB}^* and s_{HB}^{*2} , similar to the ones given by (2.4.13) and (2.4.14) for

the BHF example. The choice $a_0 = a_1 = a_2 = g_0 = g_1 = g_2 = 0$ of the hyperparameters gives a noninformative prior for the variance components. But the choice of $a_1 = 0$ or $a_2 = 0$ will give an improper posterior distribution of $(\lambda_1, \lambda_2)^T$. So we tried several combinations of these hyperparameters which are small positive numbers. Our findings for this data set, provided in Table 2.5, are not different from the BHF data set. We report our analysis for $a_0 = 0.0005$, $a_1 = 0.05$, $a_2 = 0.01$ and $g_0 = g_1 = g_2 = 0$ in Table 2.6. The estimated BLUPs for w_{13} and w_{56} reported in Harville (in press) are 10.98 and 10.29 respectively, whereas the corresponding values we obtained using a noninformative HB analysis are 11.0 and 10.4 respectively. The agreement between the two sets of estimates is remarkably close considering the fact that the underlying models (2.2.10) and (2.4.15) are not identical.

Harville (in press) also estimated the difference $w_{13} - w_{56}$ and the associated MSE of prediction by using both variance components approach and HB approach. The estimated MSE of $w_{13} - w_{56}$ in naive EBLUP approach was given by $(0.955)^2$, whereas for Kackar and Harville (1984) approximation it was $(1.053)^2$ and for Prasad and Rao (1990) approximation it was $(1.143)^2$. For HB approach, Harville (in press) used a uniform prior both for the fixed effects and the variance components associated with sire effect

Table 2.5 Birth Weights (in pounds) of Lambs

Sire	Dam	Age	Weight	Sire	Dam	Age	Weight
<u>Line 1</u>				<u>Line 4</u>			
1	1		6.2	1	1		9.2
2	1		13.0				10.6
3	1		9.5				10.6
			10.1		3		7.7
			11.4				10.0
	2		11.8				11.2
	3		12.9	2	1		10.2
			13.1				10.9
4	1		10.4	3	1		11.7
	2		8.5		3		9.9
<u>Line 2</u>				<u>Line 5</u>			
1	3		13.5	1	1		11.7
2	2		10.1				12.6
	3		11.0	2	1		9.0
			14.0		3		11.0
			15.5	3	3		9.0
3	1		12.0				12.0
4	1		11.5	4	3		9.9
	3		10.8	5	2		13.5
<u>Line 3</u>				6	2		10.9
1	2		9.0		3		5.9
	3		9.5	7	2		10.0
			12.6				12.7
2	1		10.0		3		13.2
	2		10.1				13.3
			11.7	8	1		10.7
	3		8.5				11.0
			8.8				12.5
			9.9		3		9.0
			10.9				10.2
			11.0				
			13.9				
3	1		11.6				
	3		13.0				
4	2		12.0				

Table 2.6 Predicted Birth Weights of Lambs and Associated Standard Errors

$$a_0 = .0005 \quad g_0 = 0 \quad a_1 = .05 \quad g_1 = 0 \quad a_2 = .01 \quad g_2 = 0$$

Line	Sire	e_{HB}^*	s_{HB}^*
1	1	10.1	0.90
	2	10.9	0.86
	3	11.0	0.62
	4	10.4	0.80
2	1	12.1	0.88
	2	12.2	0.71
	3	11.9	0.87
	4	11.7	0.80
3	1	10.8	0.70
	2	10.8	0.53
	3	11.3	0.75
	4	11.1	0.82
4	1	10.2	0.62
	2	10.5	0.80
	3	10.5	0.79
5	1	11.2	0.70
	2	10.7	0.70
	3	10.8	0.70
	4	10.8	0.76
	5	11.3	0.77
	6	10.4	0.74
	7	11.4	0.65
	8	10.8	0.59

and error. The HB estimate of $w_{13} - w_{56}$ in this case is reported as 0.69 and the posterior s.d. is reported as 1.042. The corresponding values obtained by using our approach are 0.60 and 0.99 respectively.

To conclude this section, we can recommend from whatever we have learned from the analysis of these data sets that the noninformative HB method is clearly a viable alternative to the usual EB or variance components approach, and should be given every serious consideration for prediction both in finite population sampling and in the infinite population situation.

2.5 Hierarchical Bayes Prediction of Finite Population Mean Vector in Absence of Unit Level Observations

Sometimes it is either difficult or impossible to obtain information at the unit level for the small areas. In this section we will derive the predictor of finite population mean vector when we do not have observations in the unit level. For $i = 1, \dots, m$, the i^{th} small area with N_i units, we assume that based on a sample of size n_i we know only the sample mean $\bar{y}_{i(s)}$ of the characteristic of interest, the sample mean vector $\bar{x}_{i(s)}$ ($p \times 1$) of the auxiliary variables. Also we have information on the population mean vector $\bar{x}_{i(p)}$ ($p \times 1$) of the auxiliary variables of the N_i units in the i^{th} small area of the population. We are interested in predicting the finite

population mean vector $(\bar{Y}_1, \dots, \bar{Y}_m)^T = \gamma$ where $\gamma_i = \bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$ based on $(\bar{y}_{1(s)}, \dots, \bar{y}_{m(s)})^T = \bar{y}_{(s)}$ (say), $(\bar{x}_{1(s)}, \dots, \bar{x}_{m(s)})^T$ and $(\bar{x}_{1(p)}, \dots, \bar{x}_{m(p)})^T$.

Let $Y_i = (Y_{i1}, \dots, Y_{iN_i})^T$, $i = 1, \dots, m$. Consider the following model.

- (i) Conditional on η_i ($N_i \times 1$), b ($p \times 1$) and δ
 $Y_i \sim N(\eta_i, R_i^{-1} I_{N_i})$, $i = 1, \dots, m$ independently,
 where R_i are known sampling variances.
- (ii) Conditional on b and δ , $\eta_i \sim N(X_i b, \delta^{-1} I_{N_i})$,
 $i = 1, \dots, m$ independently.
- (iii) b and Δ are independent a priori with
 $b \sim \text{uniform}(R^p)$ and $\Delta \sim \text{gamma}(\frac{1}{2}a, \frac{1}{2}g)$.

Combining (i) and (ii) we have

- (i') conditional on b and δ , $Y_i \sim N(X_i b, (R_i^{-1} + \delta^{-1}) I_{N_i})$,
 $i = 1, \dots, m$ independently.

Carter and Rolph (1974) introduced this type of model and Fay and Herriot (1979) considered an EB approach to this problem in a special case $\eta_i = \theta_i 1_{N_i}$ and assumed in place of (ii) that conditional on b and δ , $\theta_i \sim N(x_i^T b, \delta^{-1})$, $i = 1, \dots, m$ independently. Subsequently, in place of (i') they assumed that conditional on b and δ , $Y_i \sim N((x_i^T b) 1_{N_i}, R_i^{-1} I_{N_i} + \delta^{-1} J_{N_i})$, $i = 1, \dots, m$ independently. Fay and Herriot (1979) put a $\text{uniform}(R^p)$ prior on b . However, instead of assigning a distribution on Δ , the prior

variance, they estimated it iteratively by applying generalized least squares procedure to $\bar{Y}_{i(s)} \sim N(\mathbf{x}_i^T \mathbf{b}, R_i^{-1} n_i^{-1} + \delta^{-1})$, $i = 1, \dots, m$ independently where $\bar{Y}_{i(s)}$ is the sample mean based on n_i units. They estimated $E(\bar{Y}_i) = \theta_i$, $i = 1, \dots, m$ based on their superpopulation model whereas we are interested in predicting the finite population mean $\bar{Y}_i$, $i = 1, \dots, m$ based on (i') and (iii).

Now we will get back to our problem. For the sake of notational simplicity, we will assume without any loss of generality that the sample mean $\bar{y}_{i(s)}$ is based on the first

n_i units and is given by $\bar{y}_{i(s)} = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$. Now define

$$\bar{Y}_{i(u)} = (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} Y_{ij}, \quad \bar{\mathbf{x}}_{i(s)}^T = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}^T, \quad \bar{\mathbf{x}}_{i(u)}^T =$$

$$(N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} \mathbf{x}_{ij}^T. \quad \text{We have also defined in Section 2.4}$$

that $f_i = (N_i - n_i)/N_i$, $i = 1, \dots, m$. Since $\bar{Y}_i =$

$(1 - f_i)\bar{Y}_{i(s)} + f_i\bar{Y}_{i(u)}$ and $\bar{y}_{i(s)}$ is known, to predict the

vector $\underline{\gamma}$, it is enough to find the predictor of $(\bar{Y}_{1(u)}, \dots,$

$\bar{Y}_{m(u)})^T = \underline{\alpha}$ (say). So to predict the vector $\underline{\alpha}$, it is

enough to find the predictive distribution of $\underline{\alpha}$ given the sample mean vector $\bar{\mathbf{y}}_{(s)}$ (say).

For any quadratic loss, the predictor of γ_i is given by

$$E(\gamma_i | \bar{\mathbf{y}}_{(s)}) = (1 - f_i)\bar{y}_{i(s)} + f_i E(\alpha_i | \bar{\mathbf{y}}_{(s)}) \quad (2.5.1)$$

and its posterior variance is given by

$$V(\gamma_i | \bar{y}_{(s)}) = f_i^2 V(\alpha_i | \bar{y}_{(s)}). \quad (2.5.2)$$

Now from (i'), given b and δ , $(\bar{Y}_{1(s)}, \dots, \bar{Y}_{m(s)}, \bar{Y}_{1(u)}, \dots, \bar{Y}_{m(u)})^T$ is multivariate normal (MVN) with mean ξ and variance $\text{Diag}(\sigma_1^2, \dots, \sigma_{2m}^2)$ where $\xi_i = \bar{x}_{i(s)}^T b$, $\xi_{i+m} = \bar{x}_{i(u)}^T b$, $\sigma_i^2 = (R_i^{-1} + \delta^{-1})/n_i$ and $\sigma_{i+m}^2 = (R_i^{-1} + \delta^{-1})/(N_i - n_i)$, $i = 1, \dots, m$. From this, it is easy to derive that given $\bar{y}_{(s)}$, b and δ , α is MVN with

$$E(\alpha_i | \bar{y}_{(s)}, b, \delta) = \bar{x}_{i(u)}^T b; \quad (2.5.3)$$

$$\text{Cov}(\alpha_i, \alpha_k | \bar{y}_{(s)}, b, \delta) = \sigma_{i+m}^2 \delta_{ik} \quad (2.5.4)$$

where δ_{ik} is the Kronecker delta which is 1 if $i = k$ and is zero otherwise.

Using the iterative formulas for expectation and variance we have from (2.5.1) - (2.5.4)

$$\begin{aligned} E(\gamma_i | \bar{y}_{(s)}) &= (1 - f_i) \bar{y}_{i(s)} + f_i E \left[E(\alpha_i | \bar{y}_{(s)}, b, \delta) | \bar{y}_{(s)} \right] \\ &= (1 - f_i) \bar{y}_{i(s)} + f_i E(\bar{x}_{i(u)}^T b | \bar{y}_{(s)}) \\ &= (1 - f_i) \bar{y}_{i(s)} + f_i \bar{x}_{i(u)}^T E(b | \bar{y}_{(s)}) \end{aligned} \quad (2.5.5)$$

and

$$\begin{aligned}
V(\gamma_i | \bar{y}_{(s)}) &= f_i^2 V(\alpha_i | \bar{y}_{(s)}) \\
&= f_i^2 \left[E \left\{ V(\alpha_i | \bar{y}_{(s)}, \underline{b}, \delta) | \bar{y}_{(s)} \right\} + V \left\{ E(\alpha_i | \bar{y}_{(s)}, \underline{b}, \delta) | \bar{y}_{(s)} \right\} \right] \\
&= f_i^2 \left[E(\sigma_{i+m}^2 | \bar{y}_{(s)}) + \bar{x}_{i(u)}^T V(\underline{b} | \bar{y}_{(s)}) \bar{x}_{i(u)} \right]. \quad (2.5.6)
\end{aligned}$$

Note that from (i) and (ii), we have

(iv) given $\underline{b}$ and δ , $\bar{Y}_{(s)} \sim N(\underline{A}\underline{b}, \underline{V})$ where

$$\underline{A} = (\bar{x}_{1(s)}, \dots, \bar{x}_{m(s)})^T; \quad (2.5.7)$$

$$\underline{V} = \text{Diag}(\sigma_1^2, \dots, \sigma_m^2); \quad (2.5.8)$$

and from (iii) we can write

(v) given δ , $\underline{B} \sim \text{uniform}(R^p)$.

From (iv) and (v) we have the joint pdf of $\bar{Y}_{(s)}$ and $\underline{B}$ given δ

$$\begin{aligned}
&f(\bar{y}_{(s)}, \underline{b} | \delta) \\
&\propto \left[\prod_{i=1}^m (1/\sigma_i) \right] \exp \left[-\frac{1}{2} (\bar{y}_{(s)} - \underline{A}\underline{b})^T \underline{V}^{-1} (\bar{y}_{(s)} - \underline{A}\underline{b}) \right]. \quad (2.5.9)
\end{aligned}$$

Now assume $\text{rank}(\underline{A}) = p$ and define

$$\begin{aligned}
\hat{\underline{b}} &= (\underline{A}^T \underline{V}^{-1} \underline{A})^{-1} \underline{A}^T \underline{V}^{-1} \bar{y}_{(s)} \\
&= \left(\sum_{i=1}^m \bar{x}_{i(s)} \bar{x}_{i(s)}^T / \sigma_i^2 \right)^{-1} \left(\sum_{i=1}^m \bar{y}_{i(s)} \bar{x}_{i(s)} / \sigma_i^2 \right); \quad (2.5.10)
\end{aligned}$$

$$\underline{Q} = \underline{V}^{-1} - \underline{V}^{-1} \underline{A} (\underline{A}^T \underline{V}^{-1} \underline{A})^{-1} \underline{A}^T \underline{V}^{-1}. \quad (2.5.11)$$

With these definitions, the quadratic form in the exponent

of (2.5.9) can be written as

$$\begin{aligned} & (\bar{y}_{(s)} - A\hat{b})^T Y^{-1} (\bar{y}_{(s)} - A\hat{b}) \\ &= (\hat{b} - \hat{b})^T (A^T Y^{-1} A) (\hat{b} - \hat{b}) + \bar{y}_{(s)}^T Q \bar{y}_{(s)}. \end{aligned} \quad (2.5.12)$$

From (2.5.9) and (2.5.12), it follows that given $\bar{y}_{(s)}$ and δ , $B \sim N(\hat{b}, (A^T Y^{-1} A)^{-1})$. Note that in (2.5.10) that $\hat{b}$ depends on $\bar{y}_{i(s)}$ and δ since σ_i^2 depend on δ .

Again using the iterative formulas for expectation and variance we have

$$\begin{aligned} E(B | \bar{y}_{(s)}) &= E[E(B | \bar{y}_{(s)}, \delta) | \bar{y}_{(s)}] \\ &= E[\hat{b} | \bar{y}_{(s)}] \\ &= E\left[\left(\sum_{i=1}^m \bar{x}_{i(s)} \bar{x}_{i(s)}^T / \sigma_i^2\right)^{-1} \left(\sum_{i=1}^m \bar{y}_{i(s)} \bar{x}_{i(s)} / \sigma_i^2\right) \middle| \bar{y}_{(s)}\right] \end{aligned} \quad (2.5.13)$$

and

$$\begin{aligned} V(B | \bar{y}_{(s)}) &= E[V(B | \bar{y}_{(s)}, \delta) | \bar{y}_{(s)}] + V[E(B | \bar{y}_{(s)}, \delta) | \bar{y}_{(s)}] \\ &= E\left[(A^T Y^{-1} A)^{-1} \middle| \bar{y}_{(s)}\right] + V(\hat{b} | \bar{y}_{(s)}) \\ &= E\left[\left(\sum_{i=1}^m \bar{x}_{i(s)} \bar{x}_{i(s)}^T / \sigma_i^2\right)^{-1} \middle| \bar{y}_{(s)}\right] \\ &\quad + V\left[\left(\sum_{i=1}^m \bar{x}_{i(s)} \bar{x}_{i(s)}^T / \sigma_i^2\right)^{-1} \left(\sum_{i=1}^m \bar{y}_{i(s)} \bar{x}_{i(s)} / \sigma_i^2\right) \middle| \bar{y}_{(s)}\right]. \end{aligned} \quad (2.5.14)$$

So to evaluate $E(\gamma_i | \bar{y}_{(s)})$ and $V(\gamma_i | \bar{y}_{(s)})$, it follows from (2.5.5), (2.5.6), (2.5.13) and (2.5.14) that it is enough to evaluate $E(\sigma_{i+m}^2 | \bar{y}_{(s)})$ and the quantities that appear on the rhs of (2.5.13) and (2.5.14). In order to evaluate them, we need to find the conditional distribution of Δ given $\bar{y}_{(s)}$.

From (iii), (iv) and (v), the joint pdf of $\bar{Y}_{(s)}$, B and Δ is given by

$$f(\bar{y}_{(s)}, b, \delta) \\ \propto \left[\prod_{i=1}^m (1/\sigma_i) \right] \exp \left[-\frac{1}{2} (\bar{y}_{(s)} - A b)^T V^{-1} (\bar{y}_{(s)} - A b) \right] \delta^{\frac{1}{2}g-1} \exp \left(-\frac{1}{2} a \delta \right). \quad (2.5.15)$$

Using (2.5.12) and the fact that given $\bar{y}_{(s)}$ and δ , $B \sim N(\hat{b}, (A^T V^{-1} A)^{-1})$, integrating out b from (2.5.15) that the joint pdf of $\bar{Y}_{(s)}$ and Δ is given by

$$f(\bar{y}_{(s)}, \delta) \\ \propto \left[\prod_{i=1}^m (1/\sigma_i) \right] \exp \left[-\frac{1}{2} \{ a \delta + \bar{y}_{(s)}^T Q \bar{y}_{(s)} \} \right] \delta^{\frac{1}{2}g-1} |A^T V^{-1} A|^{-\frac{1}{2}}. \quad (2.5.16)$$

Since $f(\delta | \bar{y}_{(s)}) = \frac{f(\bar{y}_{(s)}, \delta)}{f(\bar{y}_{(s)})} \propto f(\bar{y}_{(s)}, \delta)$, the conditional distribution of Δ given $\bar{y}_{(s)}$ is determined by (2.5.16) up to the norming constant. Evaluations of (2.5.13) and

(2.5.14) are accomplished now by using (2.5.16) and typically some numerical integration techniques.

CHAPTER THREE
OPTIMALITY OF BAYES PREDICTORS
FOR MEANS IN A SPECIAL CASE

3.1 Introduction

In Chapter Two, a hierarchical Bayes procedure was introduced for prediction in mixed linear models, and in Section 2.4 the results were utilized for prediction purpose both in finite population sampling and infinite population set up in the presence of auxiliary information. There we considered the general case of unknown variance components and derived the posterior distributions of interest by assigning independent uniform prior to the fixed effects and gamma priors to the inverse of the variance components.

In this chapter, we will consider a special case. We assume that the ratios of variance components are known. We derive HB predictors for the mean vector and prove some optimal properties of this predictor.

In Section 3.2, we consider the normal linear model (2.2.2) of Section 2.2 with λ , the vector of ratios of variance components, known. We assign a uniform prior to the vector of fixed effects b and an independent gamma prior to the inverse of error variance. We first find the

posterior distribution of nonsampled units given the sampled units in finite population sampling and from this we derive the HB predictor of γ , the finite population mean vector. Later in this section, in infinite population situation, the posterior distribution of the vector of fixed and random effects and the HB predictors for linear combinations of fixed and random effects are determined. Our approach to these problems can be regarded as extensions of the HB ideas of Lindley and Smith (1972) to prediction.

Although developed within a Bayesian framework, our results should be of appeal also to frequentists. For both the problems, the BLUP notion for real valued parameters (see, for example, Henderson, 1963; Royall, 1976) is extended in Sections 3.3 and 3.4 to vector valued parameters, and it is shown that the Bayesian predictors of Section 3.2 are indeed BLUP. Like other related papers, our BLUP results do not require any normality assumption. With the added assumption of normality, the BLUPs indeed turn out to be best unbiased predictors (BUPs) within the class of all unbiased predictors. In addition, it is shown that these Bayes predictors are BUPs even for some nonnormal distributions. In these sections, we have also shown that the BLUPs also "universally" (or "stochastically") dominate (cf. Hwang, 1985) the linear unbiased predictors for elliptically symmetric

distributions. In Sections 3.5 and 3.6 we have shown that these Bayes predictors are best equivariant predictors for both the matrix loss (or standardized matrix loss) and quadratic loss (or standardized quadratic loss) under suitable groups of transformations for a broad class of elliptically symmetric distributions, including but not limited to the normal distribution.

We conclude this section by introducing a few notations. For a square matrix T (txt), $\text{tr}(T)$ denotes its trace. For a symmetric nonnegative definite (n.n.d.) matrix T , $T^{\frac{1}{2}}$ is a symmetric n.n.d. matrix such that $T^{\frac{1}{2}}T^{\frac{1}{2}} = T$, and for a symmetric p.d. matrix T , $T^{-\frac{1}{2}}$ is a symmetric p.d. matrix such that $T^{-\frac{1}{2}} = (T^{\frac{1}{2}})^{-1}$.

3.2 The Hierarchical Bayes Predictor in a Special Case

We will assume the normal linear model (2.2.2) of Section 2.2. We consider in this section the special case when λ , the vector of the ratios of variance components, is known, while B and R are independently distributed with $B \sim \text{uniform}(R^P)$ and $R \sim \text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$. Here we will consider the case of finite population sampling in details and briefly mention the corresponding results for the infinite population set up. In finite population sampling, since $\gamma = AY^{(1)} + CY^{(2)}$ for some suitable A and C , we are

still interested in finding the predictor of $\xi(Y^{(1)}, Y^{(2)})$ (recall that $\xi(Y^{(1)}, Y^{(2)}) = AY^{(1)} + CY^{(2)}$), and for this it suffices to find the predictive distribution of $Y^{(2)}$ given $Y^{(1)} = y^{(1)}$. Recall the notations K , M and G given in (2.3.4) - (2.3.6). Since λ is known in this case, we have the following Theorem 3.2.1 instead of Theorem 2.3.1. The proof of Theorem 3.2.1 is similar to that of Theorem 2.3.1 and is omitted.

Theorem 3.2.1. Assume that $n_T + g_0 - p > 2$. Then under the model given in (A) and (B) in Section 2.2 with λ known, and independent $\text{uniform}(R^p)$ prior for B and $\text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$ prior for R , the predictive distribution of $Y^{(2)}$ given $Y^{(1)} = y^{(1)}$ is multivariate t-distribution with d.f. $n_T + g_0 - p$, location parameter $My^{(1)}$ and scale parameter $(n_T + g_0 - p)^{-1} \times (a_0 + y^{(1)T}Ky^{(1)})G$.

Using the properties of the multivariate t-distribution, it is possible now to obtain closed form expressions for $E[\xi(Y^{(1)}, Y^{(2)}) | Y^{(1)} = y^{(1)}]$ and $V[\xi(Y^{(2)}, Y^{(1)}) | Y^{(1)} = y^{(1)}]$. In particular, the Bayes estimate of $\xi(Y^{(1)}, Y^{(2)})$ under any quadratic loss is now given by

$$e_{BF}^*(y^{(1)}) = (A + CM)y^{(1)}. \quad (3.2.1)$$

We may note that the predictor $\hat{e}_{BF}^*(Y^{(1)})$ given in (3.2.1) is the outcome of the model given by (A) and (B) with λ known, and the use of $\text{uniform}(R^P)$ prior on B and it does not depend on the choice of the prior (proper) distribution of R . This can be formally seen, assuming all the expectations appearing below exist, as follows:

$$\begin{aligned}
 E(Y^{(2)} | Y^{(1)}) &= E \left[E \left\{ E(Y^{(2)} | B, R, Y^{(1)}) \middle| R, Y^{(1)} \right\} \middle| Y^{(1)} \right] \\
 &= E \left[E \left\{ X^{(2)}_B + \Sigma_{21} \Sigma_{11}^{-1} (Y^{(1)} - X^{(1)}_B) \middle| R, Y^{(1)} \right\} \middle| Y^{(1)} \right] \\
 &= \Sigma_{21} \Sigma_{11}^{-1} Y^{(1)} + (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) E \left[E \left\{ B | R, Y^{(1)} \right\} \middle| Y^{(1)} \right] \\
 &= \Sigma_{21} \Sigma_{11}^{-1} Y^{(1)} + (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) \\
 &\quad \times \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} \\
 &= MY^{(1)}, \tag{3.2.2}
 \end{aligned}$$

where in the above string of equalities, the second equality follows from the fact that conditional on $B=b$, $R=r$ and $Y^{(1)} = y^{(1)}$, $Y^{(2)} \sim \left(X^{(2)}_b + \Sigma_{21} \Sigma_{11}^{-1} (y^{(1)} - X^{(1)}_b), r^{-1} \Sigma_{22.1} \right)$, the fourth follows from the fact that conditional on $R=r$ and $Y^{(1)} = y^{(1)}$, $B \sim N \left(\left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(1)T} \Sigma_{11}^{-1} Y^{(1)}, r^{-1} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \right)$ and the last equality follows from the

definition of $\underline{M}$, given in (2.3.5). Thus, the predictor $\underline{e}_{BF}^*(Y^{(1)})$ is robust against the choice of priors for R .

There are alternate ways to generate the same predictor $\underline{e}_{BF}^*(Y^{(1)})$ of $\underline{\xi}(Y^{(1)}, Y^{(2)})$. Suppose, for example, one assumes only (A) and (B) with $\underline{b}$ known (r may or may not be known). Then the best predictor (best linear predictor without the normality assumption) of $\underline{\xi}(Y^{(1)}, Y^{(2)})$ in the sense of having the smallest mean squared error matrix is given by

$$\begin{aligned} E_{\underline{\theta}} \left[\underline{\xi}(Y^{(1)}, Y^{(2)}) \middle| Y^{(1)} \right] \\ = \underline{A} Y^{(1)} + \underline{C} \left[\underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} Y^{(1)} + (X^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} X^{(1)}) \underline{b} \right] \\ \text{a.e. } (Y^{(1)}), \end{aligned} \quad (3.2.3)$$

where $\underline{\theta} = \begin{pmatrix} \underline{b} \\ r \end{pmatrix}$. (We say that $\underline{E} \leq \underline{F}$ for two symmetric matrices $\underline{E}$ and $\underline{F}$ if $\underline{F} - \underline{E}$ is n.n.d.) If $\underline{b}$ is unknown, then one replaces $\underline{b}$ by its UMVUE (BLUE without the normality assumption) $(X^{(1)T} \underline{\Sigma}_{11}^{-1} X^{(1)})^{-1} X^{(1)T} \underline{\Sigma}_{11}^{-1} Y^{(1)}$. The resulting predictor of $\underline{\xi}(Y^{(1)}, Y^{(2)})$ turns out to be $\underline{e}_{BF}^*(Y^{(1)})$. In this sense, $\underline{e}_{BF}^*(Y^{(1)})$ is also an EB predictor of $\underline{\xi}(Y^{(1)}, Y^{(2)})$.

Similarly, in this special case, one can derive the HB predictor (for quadratic loss) of $\zeta(b, y)$ in the context of infinite population set up. Denoting this HB predictor $E(\zeta(b, y)|Y)$ by $e_{BI}^*(Y)$, one can see that all the arguments leading to the empirical Bayes interpretation of $e_{BF}^*(Y^{(1)})$ work equally well to show that $e_{BI}^*(Y)$ also possesses the empirical Bayes interpretation. Harville (1985, 1988, in press) recognized this for predicting scalars.

In the next four sections, we will discuss a few frequentist properties of $e_{BF}^*(Y^{(1)})$ and $e_{BI}^*(Y)$. In Section 3.3 we show $e_{BF}^*(Y^{(1)})$ is best unbiased predictor and consider its stochastic domination whereas in Section 3.4 we consider these properties for $e_{BI}^*(Y)$. In Section 3.5 we show that $e_{BF}^*(Y^{(1)})$ is best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ under suitable groups of transformations, whereas in Section 3.6 we consider the best equivariance property of $e_{BI}^*(Y)$ under the same groups of transformations. Jeske and Harville (1987) have shown that the scalar BLUPs are best equivariant within the class of all linear equivariant predictors without any distributional assumption. However, to our knowledge, the equivariance results for vector valued predictors have not been addressed before in this context in their full generality. For an illuminating discussion on equivariant (or invariant as some authors call it) estimation, one may refer to Ferguson (1967) and Lehmann (1983).

3.3 Best Unbiased Prediction and Stochastic Domination in Small Area Estimation

In this section, we assume the normal linear model in (2.2.2) with λ known. No prior distribution for B and R is assumed, and $\underline{\theta} = (\underline{b}^T, r)^T$ is treated as an unknown parameter. First, we prove the optimality of $\underline{e}_{BF}^*(Y^{(1)})$ within the class of all unbiased predictors of $\xi(Y^{(1)}, Y^{(2)})$. Next, we dispense with the normality assumption of y and e , and prove the optimality of $\underline{e}_{BF}^*(Y^{(1)})$ within the class of all linear unbiased predictors (LUPs).

We start with the following definition of a best unbiased predictor (BUP).

Definition 3.3.1. A predictor $T(Y^{(1)})$ is said to be a BUP of $\xi(Y^{(1)}, Y^{(2)})$ if $E_{\underline{\theta}}[T(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})] = 0$ for all $\underline{\theta}$ and for every predictor $\hat{\xi}(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ satisfying $E_{\underline{\theta}}[\hat{\xi}(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})] = 0$ for all $\underline{\theta}$, $V_{\underline{\theta}}[\hat{\xi}(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})] - V_{\underline{\theta}}[T(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})]$ is n.n.d. for all $\underline{\theta}$ provided the quantities are finite.

The following general lemma plays a key role in proving the best unbiasedness of the predictor $\underline{e}_{BF}^*(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$. The lemma concerns unbiased prediction of a general $g(Y)$ ($u \times 1$) based on $Y^{(1)}$, where $g(Y)$ is not necessarily equal to $\xi(Y^{(1)}, Y^{(2)})$ or linear in Y . We

assume that each component of $\underline{g}(Y)$ has a finite second moment. Denote by $U_{\underline{g}}$, the class of all unbiased predictors $\underline{\delta}(Y^{(1)})$ of $\underline{g}(Y)$ with each component of $\underline{\delta}(Y^{(1)})$ having a finite second moment. Also, let U_0 denote the class of all real valued statistics (i.e., functions of $Y^{(1)}$) with finite second moments having zero expectations identically in $\underline{\theta}$.

Lemma 3.3.1. A predictor $T(Y^{(1)}) \in U_{\underline{g}}$ is BUP for $\underline{g}(Y)$ if and only if

$$\text{Cov}_{\underline{\theta}}[T(Y^{(1)}) - \underline{g}(Y), m(Y^{(1)})] = 0 \quad (3.3.1)$$

for all $\underline{\theta}$ and for every $m \in U_0$.

Proof of Lemma 3.3.1. Let $T(Y^{(1)}) = \left(T_1(Y^{(1)}), \dots, T_u(Y^{(1)}) \right)^T$. If $\underline{\delta}(Y^{(1)}) = \left(\delta_1(Y^{(1)}), \dots, \delta_u(Y^{(1)}) \right)^T$ is another predictor in $U_{\underline{g}}$, then

$$\begin{aligned} & V_{\underline{\theta}}[\underline{\delta}(Y^{(1)}) - \underline{g}(Y)] \\ &= V_{\underline{\theta}}[T(Y^{(1)}) - \underline{g}(Y)] + V_{\underline{\theta}}[\underline{\delta}(Y^{(1)}) - T(Y^{(1)})] \\ &\quad + \text{Cov}_{\underline{\theta}}[T(Y^{(1)}) - \underline{g}(Y), \underline{\delta}(Y^{(1)}) - T(Y^{(1)})] \\ &\quad + \text{Cov}_{\underline{\theta}}[\underline{\delta}(Y^{(1)}) - T(Y^{(1)}), T(Y^{(1)}) - \underline{g}(Y)]. \end{aligned} \quad (3.3.2)$$

Now $T_i(Y^{(1)}) - \delta_i(Y^{(1)}) \in U_0$ for every $i = 1, \dots, u$. Hence, using the condition of the lemma, for $1 \leq i \leq u$

$$\text{Cov}_{\underline{\theta}} \left[\underline{T}(\underline{Y}^{(1)}) - \underline{g}(\underline{Y}), \underline{T}_i(\underline{Y}^{(1)}) - \delta_i(\underline{Y}^{(1)}) \right] = 0. \quad (3.3.3)$$

From (3.3.2) and (3.3.3) it follows that

$$\begin{aligned} & \text{V}_{\underline{\theta}} \left[\underline{\delta}(\underline{Y}^{(1)}) - \underline{g}(\underline{Y}) \right] \\ &= \text{V}_{\underline{\theta}} \left[\underline{T}(\underline{Y}^{(1)}) - \underline{g}(\underline{Y}) \right] + \text{V}_{\underline{\theta}} \left[\underline{\delta}(\underline{Y}^{(1)}) - \underline{T}(\underline{Y}^{(1)}) \right], \end{aligned} \quad (3.3.4)$$

for all $\underline{\theta}$. Hence $\underline{T}(\underline{Y}^{(1)})$ is BUP for $\underline{g}(\underline{Y})$.

Only if. Given that $\underline{T}(\underline{Y}^{(1)})$ is BUP, we will show that the condition (3.3.1) is true. First we will show that $\underline{T}_i(\underline{Y}^{(1)})$ is BUP for $\underline{g}_i(\underline{Y})$ for every $i = 1, \dots, u$. Let $\underline{U}_i(\underline{Y}^{(1)})$ be any unbiased predictor for $\underline{g}_i(\underline{Y})$. Then $\underline{\delta}^*(\underline{Y}^{(1)})$, a u -component column vector with i^{th} component equal to $\underline{U}_i(\underline{Y}^{(1)})$, belongs to $\underline{U}_{\underline{g}}$. Then

$$\text{V}_{\underline{\theta}} \left[\underline{\delta}^*(\underline{Y}^{(1)}) - \underline{g}(\underline{Y}) \right] - \text{V}_{\underline{\theta}} \left[\underline{T}(\underline{Y}^{(1)}) - \underline{g}(\underline{Y}) \right] \text{ is n.n.d.}$$

So we have

$$\text{V}_{\underline{\theta}} \left[\underline{U}_i(\underline{Y}^{(1)}) - \underline{g}_i(\underline{Y}) \right] - \text{V}_{\underline{\theta}} \left[\underline{T}_i(\underline{Y}^{(1)}) - \underline{g}_i(\underline{Y}) \right] \geq 0, \\ i = 1, \dots, u$$

and consequently $\underline{T}_i(\underline{Y}^{(1)})$ is BUP for $\underline{g}_i(\underline{Y})$. Now following the usual Lehmann-Scheffe (1950) technique (also Rao, 1952), for any $m \in \underline{U}_0$

$$\text{Cov}_{\underline{\theta}} \left[\underline{T}_i(\underline{Y}^{(1)}) - \underline{g}_i(\underline{Y}), m(\underline{Y}^{(1)}) \right] = 0 \quad (3.3.5)$$

for all $1 \leq i \leq u$. Hence, (3.3.1) holds, and the proof of the lemma is complete.

Remark 3.3.1. It follows from the above lemma (see (3.3.4)) that if $T_1(Y^{(1)})$ and $T_2(Y^{(1)})$ are both BUPs of $g(Y)$, then

$$\begin{aligned} & E_{\theta} \left[\left(T_1(Y^{(1)}) - T_2(Y^{(1)}) \right) \left(T_1(Y^{(1)}) - T_2(Y^{(1)}) \right)^T \right] \\ &= \text{Cov}_{\theta} \left[T_1(Y^{(1)}) - g(Y), T_1(Y^{(1)}) - T_2(Y^{(1)}) \right] \\ &\quad - \text{Cov}_{\theta} \left[T_2(Y^{(1)}) - g(Y), T_1(Y^{(1)}) - T_2(Y^{(1)}) \right] \\ &= 0 - 0 = 0 \text{ by (3.3.1), for all } \theta. \end{aligned}$$

i.e., $P_{\theta} [T_1(Y^{(1)}) = T_2(Y^{(1)})] = 1$ for all θ .

Remark 3.3.2. It is also clear that the technique of the above lemma can be applied in more general contexts.

We will use the above lemma to prove the BUP property of $\xi_{BF}^*(Y^{(1)})$ in the following theorem. Recall from (3.2.1) that $\xi_{BF}^*(Y^{(1)}) = (A + CM)Y^{(1)}$.

Theorem 3.3.1. Under the normal linear model (2.2.2), $\xi_{BF}^*(Y^{(1)})$ is the BUP of $\xi(Y^{(1)}, Y^{(2)})$.

Proof of Theorem 3.3.1. In view of Lemma 3.3.1, it

suffices to show that for every $m(Y^{(1)}) \in U_0$,

$$\text{Cov}_{\theta}[\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}), m(Y^{(1)})] = 0 \text{ for all } \theta,$$

that is $E_{\theta}[\zeta(MY^{(1)} - Y^{(2)})m(Y^{(1)})] = 0$ for all θ . Since,

under the model (2.2.2), $MY^{(1)} - E_{\theta}(Y^{(2)}|Y^{(1)}) =$

$$(X^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}X^{(1)})\left[(X^{(1)T}\Sigma_{11}^{-1}X^{(1)})^{-1}X^{(1)T}\Sigma_{11}^{-1}Y^{(1)} - b\right], \text{ using}$$

$E_{\theta}[m(Y^{(1)})] = 0$, it suffices to show that

$$E_{\theta}\left[(X^{(1)T}\Sigma_{11}^{-1}Y^{(1)})m(Y^{(1)})\right] = 0 \quad (3.3.6)$$

for all θ . Since $E_{\theta}[m(Y^{(1)})] = 0$

$$\Leftrightarrow \int m(y^{(1)}) \exp\left[-\frac{1}{2}r(y^{(1)} - X^{(1)}b)^T \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}b)\right] dy^{(1)} = 0,$$

differentiating both sides of this equation w.r.t. b , one gets (see p. 318 of Rao, 1973)

$$\begin{aligned} & \int X^{(1)T}\Sigma_{11}^{-1}(y^{(1)} - X^{(1)}b)m(y^{(1)}) \\ & \quad \times \exp\left[-\frac{1}{2}r(y^{(1)} - X^{(1)}b)^T \Sigma_{11}^{-1}(y^{(1)} - X^{(1)}b)\right] dy^{(1)} \\ & = 0. \end{aligned} \quad (3.3.7)$$

Using $E_{\theta}[m(Y^{(1)})] = 0$ again, (3.3.6) follows from (3.3.7).

The proof of Theorem 3.3.1 is complete.

Remark 3.3.3. Equation (3.3.6) can be alternatively proved in the following way. Note that since $Y^{(1)} \sim N(X^{(1)}b, r^{-1}\Sigma_{11})$, $(X^{(1)T}\Sigma_{11}^{-1}Y^{(1)}, Y^{(1)T}KY^{(1)})$ is complete sufficient for θ . Hence $X^{(1)T}\Sigma_{11}^{-1}Y^{(1)}$ must have 0 covariance vector with every zero estimator $m(Y^{(1)})$, i.e.,

$$E_{\theta}\left[(X^{(1)T}\Sigma_{11}^{-1}Y^{(1)})m(Y^{(1)})\right] = 0.$$

Next we show that the conclusion of Theorem 3.3.1 continues to hold even for certain nonnormal distributions. Suppose that $e^* = (v^T, e^T)^T$ and $\Delta = \text{Diag}(\underline{D}, \Psi)$. Assume that given $R = r$, $e^* \sim N(0, r^{-1}\Delta)$, while the df of R is an arbitrary member of the family $\mathcal{F} = \{F: F \text{ is absolutely continuous with pdf } f(r) = 0 \text{ for } r \leq 0\}$. Let $\mathcal{F}^*$ denote a subfamily of $\mathcal{F}$ such that each component of $e_{BF}^*(Y^{(1)})$ and $\xi(Y^{(1)}, Y^{(2)})$ has finite second moment under the model (2.2.2) and the joint distribution of e^* and R . We now prove the following theorem.

Theorem 3.3.2. $e_{BF}^*(Y^{(1)})$ is BUP of $\xi(Y^{(1)}, Y^{(2)})$ under the model (2.2.2), $e^*|R = r \sim N(0, r^{-1}\Delta)$ and R has a df from $\mathcal{F}^*$.

Proof of Theorem 3.3.2. Using Lemma 3.3.1, and following the proof of Theorem 3.3.1, it suffices to show that

$$E_{b,F}\left[(X^{(1)T}\Sigma_{11}^{-1}Y^{(1)})m(Y^{(1)})\right] = 0 \quad (3.3.8)$$

for all $b \in R^p$ and for all $F \in \mathcal{F}^*$, where

$$E_{\underline{b}, F} \left[m(Y^{(1)}) \right] = 0 \quad (3.3.9)$$

and $E_{\underline{b}, F} \left[m^2(Y^{(1)}) \right] < \infty$ for all $\underline{b}$ and all $F \in \mathcal{F}^*$. Consider the subfamily $\mathcal{K} = \left\{ R \sim \text{gamma}\left(\frac{1}{2}c, \frac{1}{2}d\right) : c > 0, d > 2 \right\}$ of $\mathcal{F}^*$. Since (3.3.9) holds for this subfamily $\mathcal{K}$, $E_{\underline{b}, F} \left[m(Y^{(1)}) \right] = 0$ for all $\underline{b}$ and all $F \in \mathcal{K}$ gives

$$\begin{aligned} & \int_0^\infty \exp\left(-\frac{1}{2}cr\right) \int r^{\frac{1}{2}(n_T+d)-1} \\ & \quad \times \exp\left[-\frac{1}{2}r\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)^T \underline{\Sigma}_{11}^{-1}\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)\right] \\ & \quad \times m(\underline{y}^{(1)}) d\underline{y}^{(1)} dr = 0 \end{aligned} \quad (3.3.10)$$

for all $\underline{b}$, $c > 0$ and $d > 2$. Now using the uniqueness property of Laplace transforms, it follows from (3.3.10) that

$$\begin{aligned} & \int r^{\frac{1}{2}(n_T+d)-1} \exp\left[-\frac{1}{2}r\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)^T \underline{\Sigma}_{11}^{-1}\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)\right] \\ & \quad \times m(\underline{y}^{(1)}) d\underline{y}^{(1)} = 0 \end{aligned}$$

a.e. Lebesgue for all $r > 0$ and all $\underline{b}$, i.e.,

$$\begin{aligned} & \int \exp\left[-\frac{1}{2}r\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)^T \underline{\Sigma}_{11}^{-1}\left(\underline{y}^{(1)} - \underline{x}^{(1)}_{\underline{b}}\right)\right] \\ & \quad \times m(\underline{y}^{(1)}) d\underline{y}^{(1)} = 0 \end{aligned} \quad (3.3.11)$$

a.e. Lebesgue for all $r > 0$ and all $\underline{b}$. Differentiation of both sides of (3.3.11) with respect to $\underline{b}$ and some

simplifications using (3.3.11) lead to

$$\begin{aligned} & \int (\underline{x}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{y}^{(1)})_m (\underline{y}^{(1)}) \\ & \times \exp \left[-\frac{1}{2} r (\underline{y}^{(1)} - \underline{x}^{(1)} \underline{b})^T \underline{\Sigma}_{11}^{-1} (\underline{y}^{(1)} - \underline{x}^{(1)} \underline{b}) \right] \\ & \times d\underline{y}^{(1)} = 0 \end{aligned} \quad (3.3.12)$$

a.e. Lebesgue for all $r > 0$ and all $\underline{b}$. Multiplying both sides of (3.3.12) by $r^{\frac{1}{2}n_T}$ and integrating with respect to $dF(r)$ where $F \in \mathcal{F}^*$, one gets (3.3.8).

Remark 3.3.4. Since $\mathcal{F}^*$ does not contain the degenerate distributions of R on $(0, \infty)$, Theorem 3.3.1 does not follow from Theorem 3.3.2.

Remark 3.3.5. In Theorem 3.3.2, if we take $\mathcal{H}$ for $\mathcal{F}^*$, we see that the marginal distribution of $\underline{Y}$ is given by the family of distributions $\{\mathcal{T}(\cdot | N_T, \underline{x}\underline{b}, (c/d)\underline{\Sigma}, d) : \underline{b} \in \mathbb{R}^p, c > 0, d > 2\}$ and $\underline{e}_{BF}^*(\underline{Y}^{(1)})$ is BUP for $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ for this family where $\mathcal{T}(\cdot | N_T, \underline{x}\underline{b}, (c/d)\underline{\Sigma}, d)$ is N_T -variate t -distribution with location parameter $\underline{x}\underline{b}$, scale parameter $(c/d)\underline{\Sigma}$ and d.f. d .

Next we will show that the predictor $\underline{e}_{BF}^*(\underline{Y}^{(1)})$ (which is linear in $\underline{Y}^{(1)}$) is a best linear unbiased predictor (BLUP) of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. A predictor $\underline{\delta}(\underline{Y}^{(1)})$ is said to be linear if $\underline{\delta}(\underline{Y}^{(1)})$ has the form $\underline{H}\underline{Y}^{(1)}$ for some known $u \times n_T$ matrix $\underline{H}$. If in addition $E_{\theta}[\underline{\delta}(\underline{Y}^{(1)}) - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})] = 0$ for

all $\underline{\theta}$, we say that $\underline{\delta}(\underline{Y}^{(1)})$ is a LUP of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. We need the following definition.

Definition 3.3.2. A LUP $\underline{PY}^{(1)}$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ is said to be a BLUP if for every LUP $\underline{HY}^{(1)}$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$, $V_{\underline{\theta}}(\underline{HY}^{(1)} - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})) - V_{\underline{\theta}}(\underline{PY}^{(1)} - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}))$ is n.n.d. for all $\underline{\theta}$.

We now prove the BLUP property of $\underline{e}_{BF}^*(\underline{Y}^{(1)})$ for predicting $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. To this end, we will state a lemma whose proof is similar to the proof of Lemma 3.3.1 and hence the proof will be omitted.

Lemma 3.3.2. A LUP $\underline{PY}^{(1)}$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ is a BLUP if and only if

$$\text{Cov}_{\underline{\theta}}(\underline{PY}^{(1)} - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}), \underline{m}^T \underline{Y}^{(1)}) = \underline{0} \quad (3.3.13)$$

for all $\underline{\theta}$ and for every known $n_T \times 1$ vector $\underline{m}$ satisfying $E_{\underline{\theta}}(\underline{m}^T \underline{Y}^{(1)}) = 0$ for all $\underline{\theta}$.

The following theorem provides the BLUP property of $\underline{e}_{BF}^*(\underline{Y}^{(1)})$ for predicting $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. In proving this BLUP property, we do not need any distributional assumption on $\underline{e}^*$. We only assume $E_{\underline{\theta}}(\underline{e}^*) = \underline{0}$ and $V_{\underline{\theta}}(\underline{e}^*) = r^{-1} \underline{\Delta}$.

Theorem 3.3.3. Consider the linear model (2.2.2) without any distributional assumption on $\underline{e}^*$. Then $\underline{e}_{BF}^*(\underline{Y}^{(1)})$ is the BLUP of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$.

Proof of Theorem 3.3.3. If $E_{\theta}(\mathfrak{m}^T Y^{(1)}) = \mathfrak{m}^T X^{(1)} \mathfrak{b} = 0$ for all $\mathfrak{b}$, $\mathfrak{m}^T X^{(1)} = 0^T$. Hence,

$$\begin{aligned} & \text{Cov}_{\theta} \left[\varepsilon_{\text{BF}}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}), \mathfrak{m}^T Y^{(1)} \right] \\ &= \text{Cov}_{\theta} \left[\mathfrak{C}(\mathfrak{M} Y^{(1)} - Y^{(2)}), \mathfrak{m}^T Y^{(1)} \right] \\ &= \mathfrak{C}(\mathfrak{M} \Sigma_{11} - \Sigma_{21}) \mathfrak{m} \\ &= \mathfrak{C}(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} X^{(1)T} \mathfrak{m} \\ &= 0 \end{aligned}$$

for all θ . The last two equalities follow from the definition of $\mathfrak{M}$ and from the fact $\mathfrak{m}^T X^{(1)} = 0^T$. Applying Lemma 3.3.2, the result follows.

Remark 3.3.6. As already mentioned, the normality assumption is not needed for proving the BLUP property of $\varepsilon_{\text{BF}}^*(Y^{(1)})$. Theorem 3.3.3 unifies and extends the available BLUP results related to the estimation of the finite population mean vector under different models (cf. Ghosh and Lahiri, in press; Royall, 1976; and others). As in Remark 3.3.1 one can prove that the BLUP is unique with probability one.

It follows as a consequence of Theorem 3.3.3 that $\varepsilon_{\text{BF}}^*(Y^{(1)})$ has the smallest risk within the class of all LUPs of $\xi(Y^{(1)}, Y^{(2)}) \equiv \xi$ under the matrix loss

$$L_0(\xi, \underline{\xi}) = (\underline{\xi} - \xi)(\underline{\xi} - \xi)^T, \quad (3.3.14)$$

and the model (2.2.2) without any distributional assumption on $\underline{e}^*$. The optimality of $\underline{e}_{BF}^*(Y^{(1)})$ within LUPs holds a fortiori under the quadratic loss

$$\begin{aligned} L_1(\xi, \underline{\xi}) &= \|\underline{\xi} - \xi\|_{\Omega}^2 \\ &= (\underline{\xi} - \xi)^T \Omega (\underline{\xi} - \xi) \\ &= \text{tr}[\Omega L_0(\xi, \underline{\xi})], \end{aligned} \quad (3.3.15)$$

where Ω is a n.n.d. matrix. Such a loss will, henceforth, be referred to as generalized Euclidean error w.r.t. Ω . The optimality results carry over via Theorem 3.3.1 and Theorem 3.3.2 under the added distributional assumption (which is not necessarily normality assumption) on $\underline{e}^*$. A natural question to ask now is whether the risk optimality of $\underline{e}_{BF}^*(Y^{(1)})$ holds within the class of all unbiased predictors, or at least within the class of LUPs under certain other criterion for a broader family of distributions of $\underline{e}^*$. To investigate this question, we need the notions of "universal" and "stochastic" domination, and their interrelationship as given in Hwang (1985). Let $R_L(\underline{\theta}, \xi; \underline{\xi}) = E_{\underline{\theta}} \left[L \left(\left\| \underline{\xi}(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \right\|_{\Omega}^2 \right) \right]$ be the risk function of the predictor $\underline{\xi}$ for predicting ξ under a loss function which is a function of generalized Euclidean error

w.r.t. Ω for some function L . The following definition is adapted from Hwang (1985).

Definition 3.3.3. An estimator $\delta_1(Y^{(1)})$ universally dominates $\delta_2(Y^{(1)})$ (under the generalized Euclidean error w.r.t. Ω) if for every θ , and every nondecreasing loss function L , $R_L(\theta, \xi; \delta_1) \leq R_L(\theta, \xi; \delta_2)$ holds and for a particular loss, the risk functions are not identical.

Hwang (1985) has shown that (see his Theorem 2.3) δ_1 universally dominates δ_2 under the generalized Euclidean error w.r.t. Ω if $\left\| \delta_1(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \right\|_{\Omega}^2$ is stochastically smaller than $\left\| \delta_2(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \right\|_{\Omega}^2$. We say that a random variable Z_1 is stochastically smaller than Z_2 if $P_{\theta}(Z_1 > x) \leq P_{\theta}(Z_2 > x)$ for all x and θ , and for some θ , Z_1 and Z_2 have distinct distributions.

The next theorem shows that for a general class of elliptically symmetric distributions of e^* , $e_{BF}^*(Y^{(1)})$ universally dominates every LUP $H_Y^{(1)}$ of $\xi(Y^{(1)}, Y^{(2)})$ under every generalized Euclidean error w.r.t. a n.n.d. Ω . Assume that e^* has an elliptically symmetric pdf given by

$$h_*(e^* | \underline{\Delta}, r) \propto |r^{-1} \underline{\Delta}|^{-\frac{1}{2}} f(r e^{*T} \underline{\Delta}^{-1} e^*), \quad (3.3.16)$$

where as already defined $e^* = (y^T, e^T)^T$ and $\underline{\Delta} = \text{Diag}(\underline{D}, \underline{\Psi})$ and the known nonnegative function f is such that

$$\int \left[\sum_{i=1}^q |v_i| + \sum_{i=1}^{N_T} |e_i| + 1 \right] f(r \underline{e}^{*T} \underline{\Delta}^{-1} \underline{e}^*) d\underline{e}^* < \infty \quad (3.3.17)$$

where $\underline{v} = (v_1, \dots, v_q)^T$ and $\underline{e} = (e_1, \dots, e_{N_T})^T$. We will denote this distribution by $\mathfrak{E}_f(\underline{0}, r^{-1}\underline{\Delta})$ where $\mathfrak{E}_f(\underline{\mu}, \underline{\Omega}^* \sigma^2)$ denotes the distribution whose pdf is given by

$$k(\underline{t} | \underline{\mu}, \underline{\Omega}^*, \sigma) \propto |\sigma^2 \underline{\Omega}^*|^{-\frac{1}{2}} f\left((\underline{t} - \underline{\mu})^T \underline{\Omega}^{*-1} (\underline{t} - \underline{\mu}) / \sigma^2\right) \quad (3.3.18)$$

where $\underline{t}$ and $\underline{\mu}$ are in R^P , $\underline{\Omega}^*(p \times p)$ is p.d. and $\sigma > 0$.

Note that the normality of $\underline{e}^*$ with mean $\underline{0}$ and variance-covariance matrix $r^{-1}\underline{\Delta}$ is sufficient but not necessary for (3.3.16) and (3.3.17) to hold. It follows from (3.3.17) that $E(\underline{e}^*)$ exists and from (3.3.16) $E(\underline{e}^*) = \underline{0}$.

Note that $\underline{e}^{**} = (r^{-1}\underline{\Delta})^{-\frac{1}{2}} \underline{e}^*$ has a spherically symmetric distribution $\mathfrak{E}_f(\underline{0}, \underline{I}_{q^*})$ with characteristic function (c.f.) $E[\exp(i \underline{u}^T \underline{e}^{**})] = c(\underline{u}^T \underline{u})$ for some function c (see Kelker, 1970) where $i = \sqrt{-1}$, $\underline{u} = (u_1, \dots, u_{q^*})^T$ and $q^* = N_T + q$. Hence $\underline{e}^*$ has c.f. given by

$$E[\exp(i \underline{u}^T \underline{e}^*)] = c(r^{-1} \underline{u}^T \underline{\Delta} \underline{u}). \quad (3.3.19)$$

Now write $\underline{w}_j^* = \underline{z}^{(j)} \underline{v} + \underline{e}^{(j)}$ ($j = 1, 2$). Then $\underline{w}^* = (\underline{w}_1^{*T}, \underline{w}_2^{*T})^T$ has a c.f. given by

$$E[\exp(i \underline{u}^{*T} \underline{w}^*)] = c(r^{-1} \underline{u}^{*T} \underline{\Sigma} \underline{u}^*) \quad (3.3.20)$$

where $\underline{\Sigma} = \underline{\Psi} + \underline{Z}\underline{D}\underline{Z}^T$. Comparing (3.3.16) and (3.3.19) one can see from (3.3.20) that $\underline{W}^*$ has also an elliptically symmetric distribution $\mathfrak{E}_f(\underline{Q}, r^{-1}\underline{\Sigma})$ with pdf given by

$$h(\underline{w}^* | \underline{\Sigma}, r) \propto |r^{-1}\underline{\Sigma}|^{-\frac{1}{2}} f(r \underline{w}^{*T} \underline{\Sigma}^{-1} \underline{w}^*). \quad (3.3.21)$$

Theorem 3.3.4. Under the model (2.2.2), (3.3.16) and

(3.3.17), $\mathfrak{E}_{BF}^*(Y^{(1)})$ universally dominates every LUP $\mathfrak{E}(Y^{(1)}) = \mathbb{H}Y^{(1)}$ of $\xi(Y^{(1)}, Y^{(2)})$ for every p.d. $\underline{Q}$.

Remark 3.3.7. Theorem 3.3.4 does not contain Theorem 3.3.3 since Theorem 3.3.4 requires the elliptical symmetry of the distribution of $\underline{W}^*$, while the other does not. It should be noted though that the model assumption made in (3.3.16) is not necessarily stronger than the usual assumption of finiteness of certain moments. This is because the assumptions of Theorem 3.3.4 hold even if a distribution has infinite second moment (e.g., for certain multivariate t), but the BLUP property is meaningless in such instance.

Now we will state and prove a lemma. The proof of Theorem 3.3.4 rests crucially on this lemma.

Lemma 3.3.3. If $\underline{W}(N_T \times 1)$ has pdf $h(\underline{w} | \underline{I}_{N_T}, r)$, then for every $\underline{L}(u \times N_T)$, $u \leq N_T$ $\underline{L}\underline{W} \stackrel{d}{=} (\underline{L}\underline{L}^T)^{\frac{1}{2}} \underline{W}_u$, where $\underline{W}_u = (\underline{I}_u, \underline{Q})\underline{W}$ where $\underline{Q}(u \times (N_T - u))$ is a null matrix and $\stackrel{d}{=}$ means equal in distribution.

Proof of Lemma 3.3.3. The proof follows the arguments of Hwang (1985). From (3.3.19) it follows that $\underline{W}$ has c.f. $E[\exp(i\underline{t}^T \underline{W})] = c(r^{-1}\underline{t}^T \underline{t})$. Hence

$$E\left[\exp\left(i\underline{t}_1^T \underline{L}\underline{W}\right)\right] = c\left(r^{-1}\underline{t}_1^T \underline{L}\underline{L}^T \underline{t}_1\right) \quad (3.3.22)$$

where $\underline{t}_1$ is a $u \times 1$ vector. Next using (3.3.22),

$$\begin{aligned} & E\left[\exp\left(i\underline{t}_1^T (\underline{L}\underline{L}^T)^{\frac{1}{2}} \underline{W}_u\right)\right] \\ &= E\left[\exp\left(i\underline{t}_1^T (\underline{L}\underline{L}^T)^{\frac{1}{2}} (\underline{I}_u \quad 0) \underline{W}\right)\right] \\ &= c\left(r^{-1}\underline{t}_1^T (\underline{L}\underline{L}^T)^{\frac{1}{2}} (\underline{I}_u \quad 0) (\underline{I}_u \quad 0)^T (\underline{L}\underline{L}^T)^{\frac{1}{2}} \underline{t}_1\right) \\ &= c\left(r^{-1}\underline{t}_1^T (\underline{L}\underline{L}^T) \underline{t}_1\right), \end{aligned} \quad (3.3.23)$$

so that the lemma follows from (3.3.22) and (3.3.23).

It follows as a consequence of Lemma 3.3.3 that

$$\begin{aligned} \underline{W}^T \underline{L}^T \underline{L} \underline{W} &= (\underline{L}\underline{W})^T (\underline{L}\underline{W}) \\ &\triangleq \left((\underline{L}\underline{L}^T)^{\frac{1}{2}} \underline{W}_u \right)^T \left((\underline{L}\underline{L}^T)^{\frac{1}{2}} \underline{W}_u \right) = \underline{W}_u^T \underline{L}\underline{L}^T \underline{W}_u. \end{aligned} \quad (3.3.24)$$

We shall use (3.3.24) repeatedly for proving Theorem 3.3.4.

Proof of Theorem 3.3.4. Let $\underline{H}\underline{Y}^{(1)}$ be a LUP of $\xi(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. Then, under the linear model (2.2.2) and from the fact that $E(\underline{e}^*) = 0$, it is easy to see that

$$(\underline{H} - \underline{A})\underline{X}^{(1)} = \underline{C}\underline{X}^{(2)}. \quad (3.3.25)$$

Writing $\underline{W} = \underline{\Sigma}^{-\frac{1}{2}}\underline{W}^*$ and using (3.3.24) and (3.3.25) one gets

$$\begin{aligned} & \left[\underline{H}\underline{Y}^{(1)} - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) \right]^T \underline{\Omega} \left[\underline{H}\underline{Y}^{(1)} - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) \right] \\ &= \underline{W}^{*T} [\underline{H} - \underline{A} - \underline{C}]^T \underline{\Omega} [\underline{H} - \underline{A} - \underline{C}] \underline{W}^* \\ &= \underline{W}^T \underline{\Sigma}^{\frac{1}{2}} [\underline{H} - \underline{A} - \underline{C}]^T \underline{\Omega} [\underline{H} - \underline{A} - \underline{C}] \underline{\Sigma}^{\frac{1}{2}} \underline{W} \\ &\stackrel{d}{=} \underline{W}_u^T \underline{\Omega}^{\frac{1}{2}} [\underline{H} - \underline{A} - \underline{C}] \underline{\Sigma} [\underline{H} - \underline{A} - \underline{C}]^T \underline{\Omega}^{\frac{1}{2}} \underline{W}_u. \end{aligned} \quad (3.3.26)$$

Similarly,

$$\begin{aligned} & \left[\underline{e}_{BF}^*(\underline{Y}^{(1)}) - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) \right]^T \underline{\Omega} \left[\underline{e}_{BF}^*(\underline{Y}^{(1)}) - \underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)}) \right] \\ &= \underline{W}^{*T} [\underline{C}\underline{M} - \underline{C}]^T \underline{\Omega} [\underline{C}\underline{M} - \underline{C}] \underline{W}^* \\ &= \underline{W}^T \underline{\Sigma}^{\frac{1}{2}} [\underline{C}\underline{M} - \underline{C}]^T \underline{\Omega} [\underline{C}\underline{M} - \underline{C}] \underline{\Sigma}^{\frac{1}{2}} \underline{W} \\ &\stackrel{d}{=} \underline{W}_u^T \underline{\Omega}^{\frac{1}{2}} [\underline{C}\underline{M} - \underline{C}] \underline{\Sigma} [\underline{C}\underline{M} - \underline{C}]^T \underline{\Omega}^{\frac{1}{2}} \underline{W}_u. \end{aligned} \quad (3.3.27)$$

Write $\underline{\Gamma} = \underline{H} - \underline{A} - \underline{C}\underline{M}$. Then,

$$\begin{aligned} \text{rhs of (3.3.26)} &= \text{rhs of (3.3.27)} + \underline{W}_u^T \underline{\Omega}^{\frac{1}{2}} \underline{\Gamma} \underline{\Sigma}_{11} \underline{\Gamma}^T \underline{\Omega}^{\frac{1}{2}} \underline{W}_u, \\ &\quad (3.3.28) \end{aligned}$$

since using the definition of $\underline{M}$ given in (2.3.5),

$$\begin{aligned}
(\underline{C}\underline{M} - \underline{C})\underline{\Sigma}\begin{pmatrix} \underline{\Gamma}^T \\ 0 \end{pmatrix} &= \underline{C}(\underline{M} - \underline{I})\begin{pmatrix} \underline{\Sigma}_{11} \\ \underline{\Sigma}_{21} \end{pmatrix}\underline{\Gamma}^T \\
&= \underline{C}(\underline{M}\underline{\Sigma}_{11} - \underline{\Sigma}_{21})\underline{\Gamma}^T \\
&= \underline{C}(\underline{X}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})\left(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)}\right)^{-1}\underline{X}^{(1)T}\underline{\Gamma}^T,
\end{aligned} \tag{3.3.29}$$

and using (3.3.25),

$$\underline{\Gamma}\underline{X}^{(1)} = (\underline{H} - \underline{A} - \underline{C}\underline{M})\underline{X}^{(1)} = \underline{C}(\underline{X}^{(2)} - \underline{M}\underline{X}^{(1)}) = 0. \tag{3.3.30}$$

Theorem 3.3.4 follows now from (3.3.26) - (3.3.28).

Also, since $\underline{\Sigma}_{11}$ is positive definite, it follows from (3.3.28) that rhs of (3.3.26) = rhs of (3.3.27) if and only if $\underline{\Gamma} = 0$, that is $\underline{H} = \underline{A} + \underline{C}\underline{M}$ (cf. Hwang, 1985).

3.4 Best Unbiased Prediction and Stochastic Domination in Infinite Population

In this section we will briefly consider a few optimal properties of $e_{BI}^*(Y)$ which are similar to those of $e_{BF}^*(Y^{(1)})$ following closely Section 3.3. First, we note that $e_{BI}^*(Y)$ is optimal within the class of all unbiased predictors of $\zeta(b, y)$ under the normal linear model (2.2.1) with λ known. As in the finite population case, no prior distribution on B and R is assigned, and $\theta = (b^T, r)^T$ is treated as an unknown parameter. Next, dispensing with the distributional assumption of y and e , we note that $e_{BI}^*(Y)$ is BLUP for $\zeta(b, y)$.

We start with the following definition.

Definition 3.4.1. A predictor $\hat{\xi}(Y)$ of $\zeta(b, v)$ is said to be an unbiased predictor if $E_{\theta}[\hat{\xi}(Y) - \zeta(b, v)] = 0$ for all θ . An unbiased predictor $\bar{U}(Y)$ of $\zeta(b, v)$ is said to be a BUP if for every unbiased predictor $\hat{\xi}(Y)$ of $\zeta(b, v)$, $V_{\theta}[\hat{\xi}(Y) - \zeta(b, v)] - V_{\theta}[\bar{U}(Y) - \zeta(b, v)]$ is n.n.d. for all θ , provided the quantities exist finitely.

Recall that $\underline{W} = \begin{pmatrix} b \\ v \end{pmatrix}$. The following lemma is analogous to Lemma 3.3.1, and concerns the characterization of a BUP of $g(\underline{W})$ based on Y for some known function g where each component of g has a finite second moment.

Lemma 3.4.1. An unbiased predictor $\bar{U}(Y)$ of $g(\underline{W})$ with $E_{\theta}[\bar{U}^T(Y)\bar{U}(Y)] < \infty$ is BUP for $g(\underline{W})$ if and only if $\text{Cov}_{\theta}[\bar{U}(Y) - g(\underline{W}), m(Y)] = 0$ for all θ and for every statistic $m(Y)$ such that $E_{\theta}(m(Y)) \equiv 0$ and $E_{\theta}[m^2(Y)] < \infty$ for all θ .

Lemma 3.4.1 can be proved similarly as Lemma 3.3.1 and proof is omitted. We will use this lemma to sketch a proof of the following theorem which concerns best unbiased prediction of $\zeta(b, v)$.

Theorem 3.4.1. Under the normal linear model (2.2.1), $\hat{e}_{BI}^*(Y)$ is the BUP of $\zeta(b, v)$.

Proof of Theorem 3.4.1. In view of Lemma 3.4.1, and the fact that $E_{\theta}[m(Y)] = 0$ for all θ , it suffices to show that

$$E_{\theta} \left[\left\{ e_{BI}^*(Y) - \zeta(b, y) \right\} m(Y) \right] = 0 \quad \text{for all } \theta. \quad (3.4.1)$$

Note, however that with P_{θ} -probability 1

$$\begin{aligned} E_{\theta} \left[e_{BI}^*(Y) - \zeta(b, y) | Y \right] \\ &= e_{BI}^*(Y) - Sb - TDZ^T \Sigma^{-1} (Y - Xb) \\ &= \left[S(X^T \Sigma^{-1} X)^{-1} - TDZ^T \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} \right] X^T \Sigma^{-1} (Y - Xb). \end{aligned} \quad (3.4.2)$$

From (3.4.1) and (3.4.2) it suffices to show that

$$E_{\theta} \left[\left(X^T \Sigma^{-1} Y \right) m(Y) \right] = 0 \quad \text{for all } \theta.$$

This is proved similar to (3.3.6).

Remark 3.4.1. The conclusion of the above theorem holds even for certain nonnormal distributions. As in Theorem 3.3.2, one can show that $e_{BI}^*(Y)$ is BUP of $\zeta(b, y)$ under the model (2.2.1) where $e^* | R = r \sim N(0, r^{-1} \Delta)$ and R has a df from $\mathcal{T}^*$, where e^* , Δ and $\mathcal{T}^*$ are the same as in Theorem 3.3.2.

Next, note that the predictor $e_{BI}^*(Y)$ is linear in Y and it can be proved as in Theorem 3.3.3 that it is BLUP of $\zeta(b, y)$ under linear model (2.2.1) without any distributional assumption on e^* . But we need to assume e^* has mean vector 0 and finite p.d. variance-covariance matrix.

Now we will show that $e_{BI}^*(Y)$ dominates universally all LUP of $\zeta(b, y)$ for an elliptically symmetric distribution of e^* . Consider the generalized Euclidean error loss w.r.t. a u.x.p.d. matrix Ω

$$\begin{aligned} L_1(\zeta, \delta) &= \|\delta - \zeta\|_{\Omega}^2 \\ &= (\delta - \zeta)^T \Omega (\delta - \zeta). \end{aligned} \quad (3.4.3)$$

Let $R_L(\theta, \zeta; \delta) = E_{\theta} \left[L \left(\|\delta(Y) - \zeta(b, y)\|_{\Omega}^2 \right) \right]$ be the risk function of the predictor δ for predicting ζ under a loss function which is a function of generalized Euclidean error w.r.t. Ω for some function L . The following definition is similar to Definition 3.3.3.

Definition 3.4.2. An estimator $\delta_1(Y)$ universally dominates another estimator $\delta_2(Y)$ (under the generalized Euclidean error w.r.t. Ω) if for every θ and every nondecreasing function L , $R_L(\theta, \zeta; \delta_1) \leq R_L(\theta, \zeta; \delta_2)$ holds and for a particular loss, the risk functions are not identical.

Now we will state the following theorem on stochastic domination of $e_{BI}^*(Y)$. Its proof will be omitted because of its similarity to Theorem 3.3.4.

Theorem 3.4.2. Under the model (2.2.1) and an elliptically symmetric distribution of e^* similar to the ones given in

(3.3.16) and (3.3.17), $e_{BI}^*(Y)$ universally dominates every LUP $\xi(Y) = HY$ of $\zeta(b, v)$ for every p.d. Ω .

Remark 3.4.2. The version of Remark 3.3.7 in the present context is that neither the BLUP property of $e_{BI}^*(Y)$ nor the stochastic domination of $e_{BI}^*(Y)$ as given by Theorem 3.4.2 imply each other. The latter requires elliptic symmetry of distributions, while the former does not. On the other hand, the assumptions on e^* needed are similar to the ones given by (3.3.16) and (3.3.17), and do not necessarily imply finiteness of the second moments of the components of e^* .

3.5 Best Equivariant Prediction in Small Area Estimation

This section is devoted to equivariant prediction of $\xi(Y^{(1)}, Y^{(2)})$ on the basis of $Y^{(1)}$ under suitable groups of transformations. First we will assume the normal linear model given in (2.2.1). To motivate the first group of transformations as well as equivariant prediction, first consider an estimation problem. Under the normal linear model, $Y \sim N(Xb, r^{-1}\Sigma)$ where we may recall that $\Sigma = \Psi + ZDZ^T$ is known. Before proceeding further we will define $\sigma^2 = r^{-1}$ and redefine $\theta = (\sigma, b^T)^T$. We are interested in equivariant estimation of $AX^{(1)}b + CX^{(2)}b$ where the matrices $A, C, X^{(1)}$ and $X^{(2)}$ are as introduced earlier. It suffices to estimate the vector b where Xb can

be viewed as a location vector. Note that $\underline{Y} + \underline{X}\underline{\beta} \sim N(\underline{X}(\underline{b} + \underline{\beta}), \sigma^2 \underline{\Sigma})$ and it has new location vector $\underline{X}(\underline{b} + \underline{\beta})$. We are interested in developing an estimator of $\underline{b}$ which remains equivariant under a new origin of measurement. So a "natural" group of transformations will be

$$\mathcal{G}_1 = \{g_{\underline{\beta}}, \underline{\beta} \in R^p: g_{\underline{\beta}}(\underline{y}) = \underline{y} + \underline{X}\underline{\beta}\}. \quad (3.5.1)$$

Now assume that we partition $\underline{y}$, $\underline{X}$ as in (2.2.2) and only $\underline{y}^{(1)}$ is observed (which is the case in small area estimation). If $\hat{\underline{y}}^{(1)}$ estimates $(\underline{A}\underline{X}^{(1)} + \underline{C}\underline{X}^{(2)})\underline{b}$, then $\hat{\underline{y}}^{(1)} + \underline{A}\underline{X}^{(1)}\underline{\beta} + \underline{C}\underline{X}^{(2)}\underline{\beta}$ should estimate $(\underline{A}\underline{X}^{(1)} + \underline{C}\underline{X}^{(2)})(\underline{b} + \underline{\beta}) = (\underline{A} \quad \underline{C})\underline{X}(\underline{b} + \underline{\beta})$. Treating $\underline{X}(\underline{b} + \underline{\beta})$ as the new location parameter one can expect that $\hat{\underline{y}}^{(1)} + \underline{X}^{(1)}\underline{\beta}$ (estimator based on the "observed part" of $\underline{y} + \underline{X}\underline{\beta}$) will estimate $(\underline{A}\underline{X}^{(1)} + \underline{C}\underline{X}^{(2)})(\underline{b} + \underline{\beta})$. So we should have

$$\hat{\underline{y}}^{(1)} + \underline{X}^{(1)}\underline{\beta} = \hat{\underline{y}}^{(1)} + \underline{A}\underline{X}^{(1)}\underline{\beta} + \underline{C}\underline{X}^{(2)}\underline{\beta} \quad (3.5.2)$$

for all $\underline{y}^{(1)}$, $\underline{\beta}$. Now if we are interested in estimating $\xi(\underline{Y}^{(1)}, \underline{Y}^{(2)})$, instead of $E_{\underline{\theta}}(\xi(\underline{Y}^{(1)}, \underline{Y}^{(2)})) = \underline{A}\underline{X}^{(1)}\underline{b} + \underline{C}\underline{X}^{(2)}\underline{b}$, we can still use $\hat{\underline{y}}^{(1)}$ and again we will impose (3.5.2) on $\hat{\underline{y}}$.

Note that the induced group of transformations on the parameter space

$$\Theta = \{\underline{\theta}: \underline{\theta} = (\sigma, \underline{b}^T)^T, \underline{b} \in \mathbb{R}^P, \sigma > 0\} \quad (3.5.3)$$

is given by

$$\bar{\mathfrak{g}}_1 = \{\bar{\mathfrak{g}}_{\underline{\beta}}, \underline{\beta} \in \mathbb{R}^P: \bar{\mathfrak{g}}_{\underline{\beta}}(\underline{\theta}) = (\sigma, \underline{b}^T + \underline{\beta}^T)^T\}. \quad (3.5.4)$$

A loss function $L(\underline{\xi}, \underline{\theta}; \underline{\delta})$ is said to be invariant if

$$\begin{aligned} L\left(\underline{\xi}(\underline{y}^{(1)}, \underline{y}^{(2)}) + (\underline{A}\underline{X}^{(1)} + \underline{C}\underline{X}^{(2)})\underline{\beta}, \bar{\mathfrak{g}}_{\underline{\beta}}(\underline{\theta}); \right. \\ \left. \underline{\delta}(\underline{y}^{(1)}) + (\underline{A}\underline{X}^{(1)} + \underline{C}\underline{X}^{(2)})\underline{\beta}\right) \\ = L\left(\underline{\xi}(\underline{y}^{(1)}, \underline{y}^{(2)}), \underline{\theta}; \underline{\delta}(\underline{y}^{(1)})\right) \end{aligned} \quad (3.5.5)$$

for all $\underline{y}^{(1)}, \underline{y}^{(2)}, \underline{\beta}$ and $\underline{\theta}$. An estimator $\underline{\delta}$ satisfying (3.5.2) is said to be equivariant.

We will now be interested in the best equivariant prediction of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. The following lemma provides a useful characterization of the class of equivariant predictors of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. Its proof is standard and is omitted.

Lemma 3.5.1. Let $\underline{\delta}_0(\underline{Y}^{(1)})$ be an equivariant predictor for $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$. Then a necessary and sufficient condition for a predictor $\underline{\delta}(\underline{Y}^{(1)})$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ to be equivariant is that

$$\underline{\delta}(\underline{y}^{(1)}) = \underline{\delta}_0(\underline{y}^{(1)}) + \underline{h}(\underline{y}^{(1)}) \quad (3.5.6)$$

for all $\underline{y}^{(1)}$, where

$$h(\underline{y}^{(1)} + \underline{x}^{(1)}\underline{\beta}) = h(\underline{y}^{(1)}) \quad (3.5.7)$$

for all $\underline{y}^{(1)}$ and all $\underline{\beta}$.

A function h satisfying (3.5.7) is said to be invariant and hence must be a function of a maximal invariant function (see, for example, p. 285 of Lehmann, 1986) under the group of transformations

$$\mathcal{G}'_1 = \{ \underline{g}'_{\underline{\beta}}, \underline{\beta} \in \mathbb{R}^p: \underline{g}'_{\underline{\beta}}(\underline{y}^{(1)}) = \underline{y}^{(1)} + \underline{x}^{(1)}\underline{\beta} \} \quad (3.5.8)$$

induced by $\mathcal{G}_1$ in the $\underline{y}^{(1)}$ -space. The next lemma finds a maximal invariant function under the group of transformations $\mathcal{G}'_1$.

Lemma 3.5.2. A maximal invariant function under the group of transformations $\mathcal{G}'_1$ is $\underline{K}\underline{y}^{(1)}$ where $\underline{K}$ is defined in (2.3.4).

Proof of Lemma 3.5.2. First we show that $\underline{K}\underline{y}^{(1)}$ is an invariant function. From the definition of $\underline{K}$, we have $\underline{K}\underline{x}^{(1)} = \underline{0}$ and hence $\underline{K}(\underline{y}^{(1)} + \underline{x}^{(1)}\underline{\beta}) = \underline{K}\underline{y}^{(1)}$ for all $\underline{y}^{(1)}$ and $\underline{\beta}$. Now we will prove the maximality. For $\underline{y}_1^{(1)}, \underline{y}_2^{(1)}$ such that $\underline{K}\underline{y}_1^{(1)} = \underline{K}\underline{y}_2^{(1)}$ we have $\underline{y}_1^{(1)} - \underline{y}_2^{(1)}$ is orthogonal to the row space of $\underline{K}$. Also $\text{rank}(\underline{K}) = n_T - p$ and $\text{rank}(\underline{x}^{(1)}) = p$. Again using $\underline{K}\underline{x}^{(1)} = \underline{0}$ it follows that the column space of $\underline{x}^{(1)}$ forms an orthocomplement of the row space of $\underline{K}$.

Hence, $y_1^{(1)} - y_2^{(1)} = x^{(1)}\beta$ for some β , i.e., $y_1^{(1)} = y_2^{(1)} + x^{(1)}\beta$ proving thereby the maximality of $Ky^{(1)}$.

Since $e_{BF}^*(Y^{(1)})$ is an equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$, using Lemmas 3.5.1 and 3.5.2 we can characterize the class of all equivariant predictors of $\xi(Y^{(1)}, Y^{(2)})$. This is given in the next lemma.

Lemma 3.5.3. Under the group of transformations $\mathcal{G}_1$ given in (3.5.1), a predictor $\delta(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ is an equivariant predictor if and only if δ has the representation

$$\delta(y^{(1)}) = e_{BF}^*(y^{(1)}) + \ell(Ky^{(1)}) \quad (3.5.9)$$

for all $y^{(1)}$, where ℓ is an n_T -variate u -component vector valued function.

Recalling the definition of an invariant loss function $L(\xi, \theta; \delta)$ given in (3.5.5), we see that both the matrix loss L_0 and the quadratic loss L_1 , given in (3.3.14) and (3.3.15) respectively, satisfy (3.5.5).

Definition 3.5.1. An equivariant predictor $\delta_0(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ is said to be a best equivariant predictor under the loss L_0 if for every other equivariant predictor $\delta(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ $E_\theta[L_0(\xi, \delta) - L_0(\xi, \delta_0)]$ is n.n.d.

Remark 3.5.1. Note that if $\delta_0(Y^{(1)})$ is the best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ under the loss L_0 , then it is so

under the loss L_1 for every n.n.d. Ω . Conversely, if $\delta_0(Y^{(1)})$ is the best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ under the loss L_1 for every n.n.d. Ω , then it is so under L_0 . We state and prove Theorem 3.5.1 which establishes the optimality of $\varepsilon_{BF}^*(Y^{(1)})$ within the class of all equivariant predictors $\delta(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ under the loss L_0 , and a fortiori under the loss L_1 for every n.n.d. Ω .

Theorem 3.5.1. Under the normal linear model given in (2.2.2), the group of transformations $\mathfrak{g}_1$ given in (3.5.1), and the loss L_0 given in (3.3.14), the best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ is given by $\varepsilon_{BF}^*(Y^{(1)})$.

Proof of Theorem 3.5.1. Note that from (3.5.9) of Lemma 3.5.3, any equivariant predictor $\delta(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ is given by $\delta(Y^{(1)}) = \varepsilon_{BF}^*(Y^{(1)}) + \ell(KY^{(1)})$. Only those δ are to be considered for which $E_\theta[L_0(\xi, \delta)]$ is finite. Note that $E_\theta[L_0(\xi, \varepsilon_{BF}^*)]$ exists finitely and hence

$$E_\theta[\ell^T(KY^{(1)})\ell(KY^{(1)})] \leq 2\text{tr}\left\{E_\theta[L_0(\xi, \varepsilon_{BF}^*)] + E_\theta[L_0(\xi, \delta)]\right\}.$$

Hence by Cauchy-Schwarz inequality $E_\theta[\ell(KY^{(1)})\ell^T(KY^{(1)})]$ is finite. Under the matrix loss L_0 ,

$$\begin{aligned}
E_{\theta}[L_0(\xi, \delta) - L_0(\xi, \varepsilon_{BF}^*)] &= E_{\theta}[\ell(KY^{(1)})\ell^T(KY^{(1)})] \\
&+ E_{\theta}\left[\left(\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})\right)\ell^T(KY^{(1)})\right] \\
&+ E_{\theta}\left[\ell(KY^{(1)})\left(\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)})\right)^T\right]. \quad (3.5.11)
\end{aligned}$$

Now we will show that the last two product terms are null matrices. Using the definitions of $\varepsilon_{BF}^*(Y^{(1)})$ and $\underline{M}$, we have

$$\begin{aligned}
E_{\theta}\left[\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \middle| Y^{(1)}\right] \\
&= \mathbb{C}\left[\underline{MY}^{(1)} - E_{\theta}(Y^{(2)} | Y^{(1)})\right] \\
&= \mathbb{C}\left[\underline{MY}^{(1)} - \underline{X}^{(2)}\underline{b} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}(Y^{(1)} - \underline{X}^{(1)}\underline{b})\right] \text{ a.e. Leb.} \\
&= \mathbb{C}\left(\underline{X}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)}\right)(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)})^{-1} \\
&\quad \times \left(\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}Y^{(1)} - \underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{X}^{(1)}\underline{b}\right). \quad (3.5.12)
\end{aligned}$$

Now $\text{Cov}_{\theta}[\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}Y^{(1)}, KY^{(1)}] = \underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{11}K^T = \underline{X}^{(1)T}K = 0$. Since $Y^{(1)}$ is multivariate normal and $\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}Y^{(1)}$ and $KY^{(1)}$ are linear functions of $Y^{(1)}$, so $\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}Y^{(1)}$ and $KY^{(1)}$ are independently distributed. Now using (3.5.12) and independence of $\underline{X}^{(1)T}\underline{\Sigma}_{11}^{-1}Y^{(1)}$ and $KY^{(1)}$, we have

$$\begin{aligned}
& E_{\underline{\theta}} \left[\left(\varepsilon_{\text{BF}}^*(Y^{(1)}) - \underline{\xi}(Y^{(1)}, Y^{(2)}) \right) \underline{\ell}^T(KY^{(1)}) \right] \\
&= E_{\underline{\theta}} \left[E_{\underline{\theta}} \left\{ \varepsilon_{\text{BF}}^*(Y^{(1)}) - \underline{\xi}(Y^{(1)}, Y^{(2)}) \middle| Y^{(1)} \right\} \underline{\ell}^T(KY^{(1)}) \right] \\
&= \underline{C} \left(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \right) \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\
&\quad \times E_{\underline{\theta}} \left[\left(X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} - X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \underline{b} \right) \underline{\ell}^T(KY^{(1)}) \right] \\
&= \underline{C} \left(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \right) \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\
&\quad \times E_{\underline{\theta}} \left(X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} - X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \underline{b} \right) E_{\underline{\theta}} \left(\underline{\ell}^T(KY^{(1)}) \right) \\
&= 0.
\end{aligned} \tag{3.5.13}$$

Now from (3.5.11) and (3.5.13),

$$E_{\underline{\theta}} \left[L_0(\underline{\xi}, \underline{\ell}) - L_0(\underline{\xi}, \varepsilon_{\text{BF}}^*) \right] = E_{\underline{\theta}} \left[\underline{\ell}^T(KY^{(1)}) \underline{\ell}^T(KY^{(1)}) \right] \geq 0, \tag{3.5.14}$$

with equality if and only if

$$P_{\underline{\theta}} \left[\underline{\ell}^T(KY^{(1)}) = 0 \right] = 1, \text{ i.e., } P_{\underline{\theta}} \left[\underline{\ell}(Y^{(1)}) = \varepsilon_{\text{BF}}^*(Y^{(1)}) \right] = 1.$$

The proof of Theorem 3.5.1 is complete.

Next, instead of $\mathfrak{G}_1$, we consider the group $\mathfrak{G}_2$ of transformations given by

$$\mathfrak{G}_2 = \{ g_{\underline{\beta}, d}, \underline{\beta} \in \mathbb{R}^p, d > 0: g_{\underline{\beta}, d}(\underline{y}) = d\underline{y} + X\underline{\beta} \}. \tag{3.5.15}$$

A predictor $\underline{\ell}(Y^{(1)})$ of $\underline{\xi}(Y^{(1)}, Y^{(2)})$ is said to be equivariant under the group $\mathfrak{G}_2$ of transformations if

$$\delta(dy^{(1)} + X^{(1)}\beta) = d\delta(y^{(1)}) + (AX^{(1)} + CX^{(2)})\beta \quad (3.5.16)$$

for all $y^{(1)}$, β and $d > 0$. Note that $e_{BF}^*(Y^{(1)})$ is again an equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ under the above criterion. Also, note that if δ is equivariant under $\mathcal{G}_2$ taking $d = 1$ in (3.5.16) it follows that it is equivariant under $\mathcal{G}_1$. So the class of equivariant predictors under $\mathcal{G}_2$ is smaller than that under $\mathcal{G}_1$. But we will prove that $e_{BF}^*(Y^{(1)})$ is best inside this smaller class under a larger family of distributions of Y which includes the normal distribution as a special case.

We will apply the group of transformations $\mathcal{G}_2$ on the family of elliptically symmetric distributions with pdf given by

$$f_{\theta}(y) \propto |\sigma^2 \Sigma|^{-\frac{1}{2}} f\left((y - Xb)^T \Sigma^{-1} (y - Xb) / \sigma^2\right), \quad (3.5.17)$$

where f is a known nonnegative function satisfying

$$\int \left[1 + (y - Xb)^T \Sigma^{-1} (y - Xb)\right] \times f\left((y - Xb)^T \Sigma^{-1} (y - Xb) / \sigma^2\right) dy < \infty. \quad (3.5.18)$$

Note that $Y \sim N(Xb, \sigma^2 \Sigma)$ is a special case satisfying (3.5.17) and (3.5.18). It is easy to see that the above group of transformations $\mathcal{G}_2$ on N_T -dimensional Euclidean space for Y induces a group of transformations

$$\bar{\mathcal{G}}_2 = \left\{ \bar{\mathcal{G}}_{\underline{\beta}, d}, \underline{\beta} \in \mathbb{R}^p, d > 0: \bar{\mathcal{G}}_{\underline{\beta}, d}(\underline{\theta}) = (d\sigma, d\underline{b}^T + \underline{\beta}^T)^T \right\} \quad (3.5.19)$$

on the parameter space Θ given in (3.5.3).

As before, a loss function $L(\xi, \underline{\theta}; \underline{\xi})$ for predicting $\xi(Y^{(1)}, Y^{(2)})$ by $\underline{\xi}(Y^{(1)})$ is invariant under the group of transformations $\mathcal{G}_2$ if

$$\begin{aligned} & L\left(d\underline{\xi}(y^{(1)}, y^{(2)}) + (AX^{(1)} + CX^{(2)})\underline{\beta}, \bar{\mathcal{G}}_{\underline{\beta}, d}(\underline{\theta}); \right. \\ & \quad \left. d\underline{\xi}(y^{(1)}) + (AX^{(1)} + CX^{(2)})\underline{\beta}\right) \\ & = L\left(\underline{\xi}(y^{(1)}, y^{(2)}), \underline{\theta}; \underline{\xi}(y^{(1)})\right) \end{aligned} \quad (3.5.20)$$

for all $y^{(1)}, y^{(2)}, \underline{\beta}, d (>0)$ and $\underline{\theta}$.

We shall consider the special losses

$$L_2(\xi, \underline{\theta}; \underline{\xi}) = L_0(\xi, \underline{\xi})/\sigma^2 \quad (3.5.21)$$

and

$$L_3(\xi, \underline{\theta}; \underline{\xi}) = L_1(\xi, \underline{\xi})/\sigma^2. \quad (3.5.22)$$

Both these losses satisfy (3.5.20).

To find the best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ in this set up, we will present two lemmas which provide a useful characterization of the class of equivariant predictors. A proof of Lemma 3.5.4 is omitted.

Lemma 3.5.4. Let $\underline{\xi}_0(Y^{(1)})$ be an equivariant predictor for $\xi(Y^{(1)}, Y^{(2)})$. Then a necessary and sufficient condition

for a predictor $\underline{\delta}(Y^{(1)})$ of $\underline{\xi}(Y^{(1)}, Y^{(2)})$ to be equivariant is that

$$\underline{\delta}(y^{(1)}) = \underline{\delta}_0(y^{(1)}) + \underline{h}(y^{(1)}) \quad (3.5.23)$$

for all $y^{(1)}$, where

$$\underline{h}(dy^{(1)} + \underline{x}^{(1)}\underline{\beta}) = d\underline{h}(y^{(1)}) \quad (3.5.24)$$

for all $y^{(1)}$, $\underline{\beta}$ and $d > 0$.

We will follow the arguments of Lemmas 2 and 3 of Datta and Ghosh (1988) to find a representation of the function $\underline{h}$ in (3.5.24). Define

$$\underline{t}(\underline{K}y^{(1)}) = \underline{K}y^{(1)} / (y^{(1)T} \underline{K}y^{(1)})^{\frac{1}{2}} I_{[y^{(1)T} \underline{K}y^{(1)} > 0]}. \quad (3.5.25)$$

Note that since $\underline{K}\Sigma_{11}\underline{K} = \underline{K}$, so $y^{(1)T} \underline{K}y^{(1)} = (\underline{K}y^{(1)})^T \Sigma_{11}(\underline{K}y^{(1)})$ and $\underline{t}$ is indeed a function of $\underline{K}y^{(1)}$. It can be shown that $\underline{t}(\underline{K}y^{(1)})$ is a maximal invariant under the group of transformations

$$\mathcal{G}'_2 = \{g'_{\underline{\beta}, d}, \underline{\beta} \in \mathbb{R}^p, d > 0: g'_{\underline{\beta}, d}(y^{(1)}) = dy^{(1)} + \underline{x}^{(1)}\underline{\beta}\} \quad (3.5.26)$$

induced by $\mathcal{G}_2$ in the $y^{(1)}$ -space.

The following lemma characterizes the class of functions $\underline{h}(y^{(1)})$ satisfying (3.5.24). We will use this lemma to characterize the class of equivariant predictors.

Lemma 3.5.5. A function $h(y^{(1)})$ (ux1) satisfies (3.5.24) if and only if h has the representation

$$h(y^{(1)}) = (y^{(1)T} K_y^{(1)})^{\frac{1}{2}} s\left(t(K_y^{(1)})\right), \quad (3.5.27)$$

where s (ux1) is an arbitrary function of $t(K_y^{(1)})$.

Proof of Lemma 3.5.5. Assume h has the representation

given by (3.5.27). Now, since $(dy^{(1)} + x^{(1)}\beta)^T x$
 $K(dy^{(1)} + x^{(1)}\beta) = d^2 y^{(1)T} K_y^{(1)}$ and since $t(K_y^{(1)})$ is a
maximal invariant under $\mathcal{G}'_2$, so $t(K(dy^{(1)} + x^{(1)}\beta)) = t(K_y^{(1)})$
and consequently

$$\begin{aligned} h(dy^{(1)} + x^{(1)}\beta) &= (d^2 y^{(1)T} K_y^{(1)})^{\frac{1}{2}} s\left(t(K_y^{(1)})\right) \\ &= dh(y^{(1)}). \end{aligned}$$

Hence (3.5.24) is satisfied.

Only if. Since h satisfies (3.5.24) for all $y^{(1)}$, β and $d > 0$, taking $d = 1$, we see that h must satisfy
 $h(y^{(1)} + x^{(1)}\beta) = h(y^{(1)})$ for all $y^{(1)}$ and β . This implies
that h must be invariant under $\mathcal{G}'_1$, and hence must be a
function of $K_y^{(1)}$. So $h(y^{(1)}) = s(K_y^{(1)})$ where s (ux1) is
an arbitrary function satisfying

$$\begin{aligned} s\left(K(dy^{(1)} + x^{(1)}\beta)\right) &= ds(K_y^{(1)}) \\ \text{i.e., } s(dK_y^{(1)}) &= ds(K_y^{(1)}) \end{aligned} \quad (3.5.28)$$

for all $d > 0$. Now taking $d = \left(\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} \right)^{-\frac{1}{2}}$ for $\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} > 0$ we have from (3.5.28)

$$\begin{aligned} \underline{h}(\underline{y}^{(1)}) &= \underline{s}(\underline{K}_{\underline{Y}}^{(1)}) \\ &= \underline{s} \left(\left(\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} \right)^{-\frac{1}{2}} \underline{K}_{\underline{Y}}^{(1)} \right) \left(\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} \right)^{\frac{1}{2}} \\ &= \underline{s} \left(\underline{t}(\underline{K}_{\underline{Y}}^{(1)}) \right) \left(\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} \right)^{\frac{1}{2}}. \end{aligned}$$

Now for $\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} = 0$, if we take $\underline{h} = \underline{0}$, then (3.5.24) is satisfied and we can represent $\underline{h}$ by (3.5.27).

Since $\underline{e}_{\text{BF}}^*(\underline{Y}^{(1)})$ is an equivariant predictor of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$, it follows from Lemma 3.5.4 and Lemma 3.5.5 that $\underline{\delta}(\underline{Y}^{(1)})$ is an equivariant predictor of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ under the group $\mathfrak{G}_2$ of transformations if and only if

$$\underline{\delta}(\underline{y}^{(1)}) = \underline{e}_{\text{BF}}^*(\underline{y}^{(1)}) + \left(\underline{y}^{(1)T} \underline{K}_{\underline{Y}}^{(1)} \right)^{\frac{1}{2}} \underline{s} \left(\underline{t}(\underline{K}_{\underline{Y}}^{(1)}) \right) \quad (3.5.29)$$

for all $\underline{y}^{(1)}$.

Definition 3.5.2. An equivariant predictor $\underline{\delta}_0(\underline{Y}^{(1)})$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ is said to be a best equivariant predictor under the loss L_2 if for every other equivariant predictor $\underline{\delta}(\underline{Y}^{(1)})$ of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ $E_{\underline{\theta}}[L_2(\underline{\xi}, \underline{\theta}; \underline{\delta}) - L_2(\underline{\xi}, \underline{\theta}; \underline{\delta}_0)]$ is n.n.d.

Remark 3.5.2. Note that if $\underline{\delta}_0(\underline{Y}^{(1)})$ is the best equivariant predictor of $\underline{\xi}(\underline{Y}^{(1)}, \underline{Y}^{(2)})$ under the loss L_2 , then it is so under the loss L_3 for every n.n.d. $\underline{\Omega}$ and vice versa.

We now establish that $e_{BF}^*(Y^{(1)})$ is best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ under the loss L_2 . The following theorem is proved.

Theorem 3.5.2. Under the model given in (2.2.2) and the family of elliptically symmetric distributions as given by (3.5.17) and (3.5.18), the group of transformations $\mathfrak{g}_2$ given in (3.5.15), and the loss L_2 given in (3.5.21), the best equivariant predictor of $\xi(Y^{(1)}, Y^{(2)})$ is given by $e_{BF}^*(Y^{(1)})$.

Proof of Theorem 3.5.2. Note that from (3.5.29) any equivariant predictor $\delta(Y^{(1)})$ of $\xi(Y^{(1)}, Y^{(2)})$ is given by $\delta(Y^{(1)}) = e_{BF}^*(Y^{(1)}) + (Y^{(1)T} K Y^{(1)})^{\frac{1}{2}} \mathfrak{s} \left(\mathfrak{t}(K Y^{(1)}) \right)$. Only those δ are considered for which $E_{\theta}(L_2(\xi, \theta; \delta))$ is finite and in that case for the corresponding $\mathfrak{s}$, we have

$E_{\theta} \left[Y^{(1)T} K Y^{(1)} \mathfrak{s} \left(\mathfrak{t}(K Y^{(1)}) \right) \mathfrak{s}^T \left(\mathfrak{t}(K Y^{(1)}) \right) \right]$ is finite. Under the matrix loss L_2 ,

$$\begin{aligned} & \sigma^2 E_{\theta} [L_2(\xi, \theta; \delta) - L_2(\xi, \theta; e_{BF}^*)] \\ &= E_{\theta} \left[Y^{(1)T} K Y^{(1)} \mathfrak{s} \left(\mathfrak{t}(K Y^{(1)}) \right) \mathfrak{s}^T \left(\mathfrak{t}(K Y^{(1)}) \right) \right] \\ &+ E_{\theta} \left[\left(e_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \right) (Y^{(1)T} K Y^{(1)})^{\frac{1}{2}} \mathfrak{s}^T \left(\mathfrak{t}(K Y^{(1)}) \right) \right] \\ &+ E_{\theta} \left[(Y^{(1)T} K Y^{(1)})^{\frac{1}{2}} \mathfrak{s} \left(\mathfrak{t}(K Y^{(1)}) \right) \left(e_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) \right)^T \right]. \end{aligned} \tag{3.5.30}$$

It is enough to prove that the last two product terms in (3.5.30) are null matrices. Proceeding as in the proof of Theorem 3.5.1 we have

$$\begin{aligned} E_{\theta} \left[\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) | Y^{(1)} \right] \\ = C \left[MY^{(1)} - E_{\theta}(Y^{(2)} | Y^{(1)}) \right]. \end{aligned} \quad (3.5.31)$$

Now from the property of elliptically symmetric distribution, it follows that

$$E_{\theta}(Y^{(2)} | Y^{(1)}) = X^{(2)}_b + \Sigma_{21} \Sigma_{11}^{-1} (Y^{(1)} - X^{(1)}_b) \text{ a.e. Lebesgue}$$

and using this, we have from (3.5.31) that

$$\begin{aligned} E_{\theta} \left[\varepsilon_{BF}^*(Y^{(1)}) - \xi(Y^{(1)}, Y^{(2)}) | Y^{(1)} \right] \\ = C (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} \\ \times (X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} - X^{(1)T} \Sigma_{11}^{-1} X^{(1)}_b). \end{aligned} \quad (3.5.32)$$

Then,

the second term of the rhs of (3.5.30)

$$\begin{aligned} &= C (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} \\ &\times E_{\theta} \left[(Y^{(1)T} K Y^{(1)})^{\frac{1}{2}} (X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} - X^{(1)T} \Sigma_{11}^{-1} X^{(1)}_b) \right. \\ &\left. \times \Sigma^T \left(\mathfrak{t}(K Y^{(1)}) \right) \right]. \end{aligned} \quad (3.5.33)$$

Now we will show the expectation of the rhs of (3.5.33) is a null matrix. Note that since $E_{\theta}(Y^{(1)T}KY^{(1)}) < \infty$ and $E_{\theta}(Y^{(1)}) = X^{(1)}b$, by Cauchy-Schwarz inequality $(Y^{(1)T}KY^{(1)})^{\frac{1}{2}} \times X^{(1)T}\Sigma_{11}^{-1}Y^{(1)}$ has finite expectation. Now using the independence of $\left((Y^{(1)T}KY^{(1)})^{\frac{1}{2}}, X^{(1)T}\Sigma_{11}^{-1}Y^{(1)}\right)$ and $t(KY^{(1)})$ which follows as a special case of Lemma B.1, we have

$$\begin{aligned}
 & E_{\theta} \left[\left((Y^{(1)T}KY^{(1)})^{\frac{1}{2}} (X^{(1)T}\Sigma_{11}^{-1}Y^{(1)} - X^{(1)T}\Sigma_{11}^{-1}X^{(1)}b) \right) \right. \\
 & \quad \left. \times \mathbb{S}^T(t(KY^{(1)})) \right] \\
 &= E_{\theta} \left[\left((Y^{(1)T}KY^{(1)})^{\frac{1}{2}} X^{(1)T}\Sigma_{11}^{-1}(Y^{(1)} - X^{(1)}b) \right) \right] \\
 & \quad \times E_{\theta} \left[\mathbb{S}^T(t(KY^{(1)})) \right]. \tag{3.5.34}
 \end{aligned}$$

Next using the elliptical symmetry of $Y^{(1)}$, it follows that

$$Y^{(1)} - X^{(1)}b \stackrel{d}{=} -(Y^{(1)} - X^{(1)}b).$$

Now since $KX^{(1)} = 0$ and K is n.n.d., one has

$$\begin{aligned}
 & \left[(Y^{(1)} - X^{(1)}b)^T K (Y^{(1)} - X^{(1)}b) \right]^{\frac{1}{2}} \\
 & \quad \times \left[X^{(1)T}\Sigma_{11}^{-1}(Y^{(1)} - X^{(1)}b) \right]
 \end{aligned}$$

$$\begin{aligned} & \stackrel{d}{=} \left[\left\{ -\left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right) \right\}^T \underline{K} \left\{ -\left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right) \right\} \right]^{\frac{1}{2}} \\ & \quad \times \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \left\{ -\left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right) \right\} \end{aligned}$$

i.e.,

$$\begin{aligned} & \left(\underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)} \right)^{\frac{1}{2}} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right) \\ & \stackrel{d}{=} -\left(\underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)} \right)^{\frac{1}{2}} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right). \end{aligned} \quad (3.5.35)$$

Now taking expectations on both sides which exist finitely due to the previous observation, one gets

$$E_{\underline{\theta}} \left[\left(\underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)} \right)^{\frac{1}{2}} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \left(\underline{Y}^{(1)} - \underline{X}^{(1)}_{\underline{b}} \right) \right] = 0. \quad (3.5.36)$$

Hence from (3.5.30), (3.5.33), (3.5.34) and (3.5.36) we have

$$\begin{aligned} & \sigma^2 E_{\underline{\theta}} \left[L_2(\underline{\xi}, \underline{\theta}; \underline{\ell}) - L_2(\underline{\xi}, \underline{\theta}; \underline{\varepsilon}_{BF}^*) \right] \\ & = E_{\underline{\theta}} \left[\left(\underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)} \right) \underline{\varepsilon} \left(\underline{t}(\underline{K} \underline{Y}^{(1)}) \right) \underline{\varepsilon}^T \left(\underline{t}(\underline{K} \underline{Y}^{(1)}) \right) \right] \end{aligned}$$

which is n.n.d. for all $\underline{\theta}$ and the two risk matrices are equal if and only if

$$P_{\underline{\theta}} \left[\underline{\varepsilon} \left(\underline{t}(\underline{K} \underline{Y}^{(1)}) \right) = 0 \right] = 1 \quad \text{for all } \underline{\theta}$$

$$\text{i.e., } P_{\underline{\theta}} \left[\underline{\ell}(\underline{Y}^{(1)}) = \underline{\varepsilon}_{BF}^*(\underline{Y}^{(1)}) \right] = 1 \text{ for all } \underline{\theta}.$$

Remark 3.5.3. Although $\left((Y^{(1)T} K Y^{(1)})^{\frac{1}{2}}, X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} \right)$ is sufficient and $t(K Y^{(1)})$ is ancillary, Basu's Theorem (1955) can not be applied since the sufficient statistic is not complete.

3.6 Best Equivariant Prediction in Infinite Population

In this section, we concentrate on the equivariant prediction of $\zeta(\underline{b}, \underline{y}) = \underline{S}\underline{b} + \underline{T}\underline{y}$ on the basis of the data vector $\underline{Y}$ under suitable groups of transformations. We will first consider the prediction of $\zeta(\underline{b}, \underline{y})$ under the normal linear model (2.2.1) and the group of transformations $\mathcal{G}_1$ given in (3.5.1). Later on we will use the broader group of transformations $\mathcal{G}_2$ and elliptically symmetric distributions for $\underline{Y}$ for equivariant prediction of $\zeta(\underline{b}, \underline{y})$. Jeske and Harville (1987) and Harville (in press) considered equivariant prediction of a scalar linear combination of $\underline{b}$ and $\underline{y}$ under the group of transformations $\mathcal{G}_1$.

As before, we say a predictor $\hat{\zeta}(\underline{Y})$ of $\zeta(\underline{b}, \underline{y})$ is equivariant under the group of transformations $\mathcal{G}_1$ if

$$\hat{\zeta}(\underline{y} + \underline{X}\underline{\beta}) = \hat{\zeta}(\underline{y}) + \underline{S}\underline{\beta}, \text{ for all } \underline{y} \text{ and } \underline{\beta}. \quad (3.6.1)$$

A loss function $L(\zeta, \underline{\theta}; \hat{\zeta})$ is said to be invariant if

$$\begin{aligned} L(\zeta(\underline{b}, \underline{y}) + \underline{S}\underline{\beta}, \underline{g}_{\underline{\beta}}(\underline{\theta}); \hat{\zeta}(\underline{y}) + \underline{S}\underline{\beta}) \\ = L(\zeta(\underline{b}, \underline{y}), \underline{\theta}; \hat{\zeta}(\underline{y})) \end{aligned} \quad (3.6.2)$$

for all $\underline{y}$, $\underline{v}$, $\underline{\beta}$ and $\underline{\theta}$. Note that the matrix loss $L_0(\underline{\zeta}, \underline{\delta}) = (\underline{\delta} - \underline{\zeta})(\underline{\delta} - \underline{\zeta})^T$ and the quadratic loss $L_1(\underline{\zeta}, \underline{\delta}) = (\underline{\delta} - \underline{\zeta})^T \underline{\Omega}(\underline{\delta} - \underline{\zeta})$ are invariant. Now we have the following lemma similar to Lemma 3.5.1 which characterizes the class of equivariant predictors of $\underline{\zeta}(\underline{b}, \underline{v})$ based on a specific equivariant predictor $\underline{\delta}_0(\underline{Y})$.

Lemma 3.6.1. A necessary and sufficient condition for a predictor $\underline{\delta}(\underline{Y})$ of $\underline{\zeta}(\underline{b}, \underline{v})$ to be equivariant is that

$$\underline{\delta}(\underline{y}) = \underline{\delta}_0(\underline{y}) + \underline{h}(\underline{y}) \quad (3.6.3)$$

for all $\underline{y}$, where $\underline{\delta}_0(\underline{Y})$ is a specific equivariant predictor and the vector valued function $\underline{h}$ satisfies

$$\underline{h}(\underline{y} + \underline{X}\underline{\beta}) = \underline{h}(\underline{y}) \quad (3.6.4)$$

for all $\underline{y}$ and $\underline{\beta}$.

The function $\underline{h}$ is said to be invariant and hence it must be a function of a maximal invariant function under $\mathcal{G}_1$. So in the following lemma we find a maximal invariant function under the group of transformations $\mathcal{G}_1$. This is only a restatement of Lemma 3.5.2. So the proof is omitted.

Lemma 3.6.2. A maximal invariant function under $\mathcal{G}_1$ is $\underline{Q}\underline{y}$ where $\underline{Q} = \underline{\Sigma}^{-1} - \underline{\Sigma}^{-1}\underline{X}(\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1}$ as defined in (2.3.12).

Since $e_{BI}^*(Y + X\beta) = e_{BI}^*(Y) + S\beta$, so $e_{BI}^*(Y)$ is an equivariant predictor of $\zeta(b, y) = Sb + Ty$. This observation and Lemmas 3.6.1 and 3.6.2 culminate into Lemma 3.6.3 which provides a useful characterization of the class of equivariant predictors under $\mathcal{G}_1$.

Lemma 3.6.3. Under the group of transformations $\mathcal{G}_1$, any equivariant predictor $\hat{\ell}(Y)$ can be represented as

$$\hat{\ell}(y) = e_{BI}^*(y) + \ell(Qy) \quad (3.6.5)$$

for all y , where ℓ is an n_T -variate u -component vector valued function.

As in Definition 3.5.1, we say an equivariant predictor $\hat{\ell}_0(Y)$ of $\zeta(b, y)$ is best under the loss L_0 if for every other equivariant predictor $\hat{\ell}(Y)$ $E_\theta[L_0(\zeta, \hat{\ell}) - L_0(\zeta, \hat{\ell}_0)]$ is n.n.d. As in Remark 3.5.1 it is true that an equivariant predictor of $\zeta(b, y)$ is best under the loss L_0 if and only if it is best under the loss L_1 for every n.n.d. matrix Ω . Similar to Theorem 3.5.1 we have the following theorem under the assumption that $e^* = \begin{pmatrix} y \\ e \end{pmatrix} \sim N(Q, \sigma^2 \Delta)$ where $\Delta = \text{Diag}(D, \Psi)$ is known p.d. and $\sigma^2 > 0$ is unknown.

Theorem 3.6.1. Under the model (2.2.1) with the distributional assumption on e^* given above, the group of

transformations $\mathfrak{G}_1$ and the matrix loss L_0 , the best equivariant predictor of $\zeta(\mathfrak{b}, \mathfrak{y})$ is given by $\mathfrak{e}_{BI}^*(Y)$.

A proof of this theorem is similar to that of Theorem 3.5.1. Evaluating first the conditional expectation $E_{\theta}[\zeta(\mathfrak{b}, \mathfrak{y})|Y] = S\mathfrak{b} + TE_{\theta}[\mathfrak{y}|Y]$, which is a linear function of Y , one can proceed as in Theorem 3.5.1 to complete the proof.

To conclude this section, we will consider the group of transformations $\mathfrak{G}_2$ and a class of elliptically symmetric distributions to show that $\mathfrak{e}_{BI}^*(Y)$ is the best equivariant predictor of $\zeta(\mathfrak{b}, \mathfrak{y})$. Assume that $\mathfrak{e}^*$ has an elliptically symmetric distribution with pdf given by

$$p(\mathfrak{e}^*|\sigma) \propto |\sigma^2 \underline{\Delta}|^{-\frac{1}{2}} f(\mathfrak{e}^{*T} \underline{\Delta}^{-1} \mathfrak{e}^* / \sigma^2) \quad (3.6.6)$$

where f is some known nonnegative function satisfying the condition

$$\int [1 + \mathfrak{e}^{*T} \underline{\Delta} \mathfrak{e}^*] f(\mathfrak{e}^{*T} \underline{\Delta}^{-1} \mathfrak{e}^* / \sigma^2) d\mathfrak{e}^* < \infty. \quad (3.6.7)$$

where $\underline{\Delta}$ and σ^2 are the same as above. Now we have the following theorem.

Theorem 3.6.2. Under the model (2.2.1) with the distributional assumptions on $\mathfrak{e}^*$ given by (3.6.6) and (3.6.7), the group of transformations $\mathfrak{G}_2$ and the standardized matrix loss $L_2(\zeta, \theta; \mathfrak{e}) = L_0(\zeta, \mathfrak{e})/\sigma^2$, the best equivariant predictor of $\zeta(\mathfrak{b}, \mathfrak{y})$ is given by $\mathfrak{e}_{BI}^*(Y)$.

To avoid the repetition of the arguments a detailed proof of this theorem is omitted. Proceeding exactly as in Theorem 3.5.2 and Theorem 3.6.1 we can prove this theorem.

CHAPTER FOUR
ASYMPTOTIC OPTIMALITY OF
HIERARCHICAL BAYES PREDICTORS FOR MEANS

4.1 Introduction

We have introduced in the previous chapters several HB predictors both in the context of finite and infinite population sampling under general linear models. There we have used uniform prior (over the appropriate Euclidean space) for the fixed effects and inverse gamma (possibly improper) for the variance components to predict $\underline{\gamma} = (\gamma_1, \dots, \gamma_m)^T$, the vector of finite population means. A natural question to ask is that if indeed there is a "true" or "elicited" prior, whether the HB predictor of $\underline{\gamma}$ is, in some sense, close to the "true" Bayes predictor which we shall refer to as subjective Bayes predictor. Such a comparison can be made conveniently in terms of Bayes risks of these predictors computed under the elicited prior. Following Robbins (1955), we shall call a predictor of $\underline{\gamma}$ "asymptotically optimal" (A.O.) if the difference in the Bayes risks of the predictor and the subjective Bayes predictor converges to zero as $m \rightarrow \infty$. The A.O. property of certain EB predictors arising naturally in the context

of finite population sampling was proved in Ghosh and Meeden (1986), Ghosh and Lahiri (1987a), and Ghosh, Lahiri and Tiwari (1989).

In this chapter, we prove the A.O. property of several HB predictors under average squared error loss. To our knowledge, such an attempt is the first of its kind. In Section 4.2, we start with the normal linear model (2.2.2) when λ , the vector of the ratios of variance components, is known. We specify our loss function, the elicited prior distribution and derive the subjective Bayes predictor of γ . In Subsection 4.2.1, we derive a general expression for the difference of Bayes risks of our subjective Bayes predictor and any predictor of γ . In particular, we consider the sample mean vector and the HB predictor of γ which can be derived from Section 3.2. In Subsection 4.2.2, we consider the random regression coefficients model of Dempster et al. (1981) and in Subsection 4.2.3 the nested error regression model of Battese et al. (1988) as special cases of the general model of Subsection 4.2.1. We have shown that the HB predictor is asymptotically optimal, whereas the traditional sample mean vector is nonoptimal.

In Section 4.3, we consider the Fay-Herriot model of Section 2.5 with all the sampling variances $V_i = R_i^{-1}$ (say) known. We assume for the sample sizes n_i , V_i/n_i are all equal. We first show the A.O. property of the HB predictor

of $\underline{\theta}$ ($m \times 1$) where $\underline{\theta}$ is as given in Section 2.5. The result is then used to prove the A.O. property of the HB predictor of $\underline{\gamma}$. The HB predictor of $\underline{\theta}$ is the shrinkage estimator obtained by shrinking the maximum likelihood estimator of $\underline{\theta}$ towards a regression surface. The model we will consider in Section 4.3 is a slight generalization of the ones given in Morris (1981, 1983) and Ghosh (1989), and includes also as special cases the ones considered in Strawderman (1971) and Faith (1978).

The assumption of known first stage variance component of Section 4.3 is dispensed with in Section 4.4, and a prior distribution (proper or improper) is assigned to the first stage variance component. This situation is now a special case of the nested error regression model where the covariate vectors associated with each unit within a stratum are the same. In Stroud (1987) (see also Ghosh and Lahiri, in press), one can find such an example. Once again, the A.O. property of HB predictor of $\underline{\theta}$ is proved. The result is then used to prove the A.O. property of the HB predictor of $\underline{\gamma}$.

In the remainder of this section, we will introduce a few notations. For a sequence of $a \times b$ matrices

$\left\{ \underline{F}(m) = \left((f_{ij}(m)) \right) \right\}$ and for an $a \times b$ matrix $\underline{F}^0 = \left((f_{ij}^0) \right)$, we

say $\lim_{m \rightarrow \infty} \underline{F}(m)$ exists and write $\lim_{m \rightarrow \infty} \underline{F}(m) = \underline{F}^0$ if

$\lim_{m \rightarrow \infty} f_{ij}(m) = f_{ij}^0$ for $i = 1, \dots, a$, $j = 1, \dots, b$. For an $a \times a$

matrix T , we denote its largest (smallest) eigen value by $ch_L(T)$ ($ch_S(T)$).

4.2 Model, Loss, Prior and Predictors

We start with the linear model (2.2.2), namely,

$$\begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \underline{b} + \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix} \underline{\gamma} + \begin{pmatrix} e^{(1)} \\ e^{(2)} \end{pmatrix} \quad (4.2.1)$$

where all the quantities appearing in (4.2.1) are the same as in (2.2.2). We want to study A.O. property of predictors of $\underline{\gamma}$, the finite population mean vector. Let

$\underline{A} = \bigoplus_{i=1}^m (N_i^{-1} \underline{1}_{n_i}^T)$ and $\underline{C} = \bigoplus_{i=1}^m (N_i^{-1} \underline{1}_{N_i - n_i}^T)$ where n_i and N_i are the same as before. Then $\underline{\gamma} = \underline{A} \underline{Y}^{(1)} + \underline{C} \underline{Y}^{(2)}$. Similarly, write the sample mean vector $\bar{Y}_{(s)} = (\bar{Y}_{1(s)}, \dots, \bar{Y}_{m(s)})^T$ as $\bar{Y}_{(s)} = \underline{L} \underline{Y}^{(1)}$ where $\bar{Y}_{i(s)} = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$, $i = 1, \dots, m$ and $\underline{L} = \bigoplus_{i=1}^m (n_i^{-1} \underline{1}_{n_i}^T)$.

Consider the quadratic loss function

$$L(\underline{\gamma}, \underline{a}) = m^{-1} (\underline{a} - \underline{\gamma})^T \underline{Q}_m (\underline{a} - \underline{\gamma}) \quad (4.2.2)$$

where $\underline{\gamma}$ is estimated by the vector $\underline{a}$ ($m \times 1$) and $\underline{Q}_m$ is a known $m \times m$ n.n.d. and nonnull matrix.

In this section we assume λ to be known, say $\lambda = \lambda_0$. We will consider the elicited prior distribution π_0 under

which $Y \sim N(Xb_0, r_0^{-1}\Sigma)$ where b_0 ($p \times 1$), r_0 (>0) and $\Sigma = \Psi + ZD(\lambda_0)Z^T$ are known.

Now the subjective Bayes predictor of γ under the loss (4.2.2) is given by

$$\begin{aligned} \epsilon_{SB}(Y^{(1)}) &= AY^{(1)} + CE\pi_0(Y^{(2)}|Y^{(1)}) \\ &= AY^{(1)} + C\left[X^{(2)}b_0 + \Sigma_{21}\Sigma_{11}^{-1}(Y^{(1)} - X^{(1)}b_0)\right] \end{aligned} \quad (4.2.3)$$

where Σ_{ij} $i, j = 1, 2$ are partitioned matrices of Σ .

For known λ , the HB predictor of γ with independent uniform(R^p) prior for B and $\text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$ prior for R and the loss (4.2.2) follows from Section 3.2 and is given by

$$\epsilon_{BF}^*(Y^{(1)}) = (A + CM)Y^{(1)} \quad (4.2.4)$$

where we can recall from (2.3.5) that

$$M = \Sigma_{21}\Sigma_{11}^{-1} + (X^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}X^{(1)})(X^{(1)T}\Sigma_{11}^{-1}X^{(1)})^{-1}X^{(1)T}\Sigma_{11}^{-1}. \quad (4.2.5)$$

The Bayes risk $r_{Q_m}(\pi_0, \delta)$ of a predictor $\delta(Y^{(1)})$ of $\gamma \equiv \gamma(Y)$ when the prior is π_0 is given by

$$\begin{aligned} r_{Q_m}(\pi_0, \delta) &= E_{\pi_0} \left[m^{-1} \left(\delta(Y^{(1)}) - \gamma(Y) \right)^T Q_m \left(\delta(Y^{(1)}) - \gamma(Y) \right) \right]. \end{aligned} \quad (4.2.6)$$

4.2.1 General Expressions for Bayes Risks Difference

We will now derive the difference in the Bayes risks of an arbitrary predictor $\hat{\gamma}(Y^{(1)})$ of γ and the subjective Bayes predictor $\epsilon_{SB}(Y^{(1)})$. For the quadratic loss (4.2.2), noting that $\epsilon_{SB}(Y^{(1)}) = E_{\pi_0}[\gamma(Y) | Y^{(1)}]$, standard Bayesian calculations give

$$\begin{aligned} r_{Q_m}(\pi_0, \hat{\gamma}) - r_{Q_m}(\pi_0, \epsilon_{SB}) \\ = m^{-1} E_{\pi_0} \left[\left(\hat{\gamma}(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right)^T Q_m \left(\hat{\gamma}(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right) \right]. \end{aligned} \quad (4.2.1.1)$$

For $\hat{\gamma}(Y^{(1)}) = \epsilon_{BF}^*(Y^{(1)})$, the rhs of (4.2.1.1) is

$$\begin{aligned} m^{-1} E_{\pi_0} \left[\left(\epsilon_{BF}^*(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right)^T Q_m \left(\epsilon_{BF}^*(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right) \right] \\ = m^{-1} \text{tr} \left[Q_m E_{\pi_0} \left(\epsilon_{BF}^*(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right) \left(\epsilon_{BF}^*(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \right)^T \right]. \end{aligned} \quad (4.2.1.2)$$

Now from (4.2.3) and (4.2.4), we have

$$\begin{aligned} \epsilon_{BF}^*(Y^{(1)}) - \epsilon_{SB}(Y^{(1)}) \\ = C \left[M Y^{(1)} - X^{(2)} b_0 - \Sigma_{21} \Sigma_{11}^{-1} (Y^{(1)} - X^{(1)} b_0) \right] \\ = C \left(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \right) \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\ \times \left(X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} - X^{(1)T} \Sigma_{11}^{-1} X^{(1)} b_0 \right). \end{aligned}$$

Using this, from (4.2.1.1) and (4.2.1.2) we obtain

$$\begin{aligned} r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \\ = r_0^{-1} m^{-1} \text{tr} \left[Q_m C (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) \right. \\ \left. \times (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T \right]. \quad (4.2.1.3) \end{aligned}$$

For $\hat{\delta}(Y^{(1)}) = \bar{Y}_{(s)}$, we have from (4.2.1.1),

$$\begin{aligned} r_{Q_m}(\pi_0, \bar{Y}_{(s)}) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \\ = \text{tr} \left[m^{-1} Q_m \left(E \pi_0(\bar{Y}_{(s)} - \varepsilon_{SB}(Y^{(1)})) E \pi_0(\bar{Y}_{(s)} - \varepsilon_{SB}(Y^{(1)}))^T \right) \right] \\ + \text{tr} \left[r_0^{-1} m^{-1} Q_m (L - A - C \Sigma_{21} \Sigma_{11}^{-1}) \Sigma_{11} (L - A - C \Sigma_{21} \Sigma_{11}^{-1})^T \right] \\ = \text{tr} \left[m^{-1} Q_m (L X^{(1)} - A X^{(1)} - C X^{(2)}) b_0 b_0^T (L X^{(1)} - A X^{(1)} - C X^{(2)})^T \right] \\ + \text{tr} \left[r_0^{-1} m^{-1} Q_m (L - A - C \Sigma_{21} \Sigma_{11}^{-1}) \Sigma_{11} (L - A - C \Sigma_{21} \Sigma_{11}^{-1})^T \right]. \quad (4.2.1.4) \end{aligned}$$

We will use (4.2.1.3) and (4.2.1.4) repeatedly in the following subsections.

4.2.2 Random Regression Coefficients Model

We will examine the behavior of the risk difference given by (4.2.1.3) and (4.2.1.4) in this special case. For simplicity, a random regression coefficients model as

appeared in Prasad and Rao (1990) will be considered. The model can be described as follows:

- (i) Conditional on $R = r$, $B = b$ and $B_i = b_i$, $i = 1, \dots, m$; $Y_{ij} \sim N(b_i x_{ij}, r^{-1})$, $j = 1, \dots, N_i$, $i = 1, \dots, m$ independently;
- (ii) conditional on $R = r$ and $B = b$, $B_i \sim N(b, (r\lambda_0)^{-1})$, $i = 1, \dots, m$ independently;
- (iii) B is uniform $(-\infty, \infty)$ and $R \sim \text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$ independently.

We can write (i) and (ii) as

$$Y_{ij} = b x_{ij} + v_i x_{ij} + e_{ij},$$

($j = 1, \dots, N_i$, $i = 1, \dots, m$) where e_{ij} and v_i are mutually independent with $v_i \sim N(0, (r\lambda_0)^{-1})$ and $e_{ij} \sim N(0, r^{-1})$.

Here $\underline{X} = \text{col}_{1 \leq i \leq m}(\underline{x}_i)$, $\underline{x}_i = (x_{i1}, \dots, x_{iN_i})^T$, $i = 1, \dots, m$, $\underline{Z} = \bigoplus_{i=1}^m \underline{x}_i$, $\underline{\Psi} = I_{N_T}$, $\underline{D}(\lambda) = \lambda_0^{-1} I_m$ and $\underline{\Sigma} = I_{N_T} + \lambda_0^{-1} \bigoplus_{i=1}^m \underline{x}_i \underline{x}_i^T$. We

will show that under appropriate conditions the risk

difference given by (4.2.1.3) goes to zero, whereas the

risk difference given by (4.2.1.4) does not go to zero as m

$\rightarrow \infty$. To this end, we will prove two theorems. The first

theorem will prove the A.O. property of the HB predictor

whereas the second theorem proves the nonoptimality of the

sample mean vector. To prove Theorem 4.2.2.1 below, we

need the following conditions. Assume

$$\overline{\lim}_{m \rightarrow \infty} \text{ch}_L(Q_m) = \Gamma \text{ (say) } < \infty, \quad (4.2.2.1)$$

and for two positive numbers p_1 and p_2 with $p_1 < p_2$

$$p_1 \leq x_{ij} \leq p_2 \text{ for } j = 1, \dots, N_i, \quad i = 1, \dots, m. \quad (4.2.2.2)$$

Before getting into the theorems, we will make a remark on the condition (4.2.2.1) which pertains to all the theorems in this chapter.

Remark 4.2.2.1. Note that for average squared error loss $Q_m = I_m$ and the condition (4.2.2.1) trivially holds.

Theorem 4.2.2.1. Suppose under the prior π_0 , $Y \sim N(Xb_0, r_0^{-1}\Sigma)$ where b_0 is a scalar. Assume the conditions (4.2.2.1) and (4.2.2.2) hold. Then for the random regression coefficients model the HB predictor $e_{BF}^*(Y^{(1)})$ in (4.2.4) of γ is asymptotically optimal under the loss (4.2.2).

To prove the theorem, we need the following lemma. We will use this lemma repeatedly in this chapter. Also, in our applications the matrix P of the lemma happens to be n.n.d.

Lemma 4.2.2.1. Let $P(p \times p)$ be a symmetric matrix and $U(q \times p)$ be an arbitrary matrix. Then

$$\text{ch}_S(P) \text{tr}(U^T U) \leq \text{tr}(P U^T U) \leq \text{ch}_L(P) \text{tr}(U^T U).$$

Proof of Lemma 4.2.2.1. Use the spectral decomposition

$P = \sum_{i=1}^p \eta_i \xi_i \xi_i^T$, where η_i are the eigen values of P and ξ_i are the corresponding orthonormal eigen vectors. Then,

$$\text{tr}(PU^TU) = \sum_{i=1}^p \eta_i \text{tr}(\xi_i \xi_i^T U^T U) = \sum_{i=1}^p \eta_i \text{tr}[(U \xi_i)(U \xi_i)^T].$$

Since $\text{ch}_S(P) \leq \eta_i \leq \text{ch}_L(P)$, $\sum_{i=1}^p \xi_i \xi_i^T = I_p$ and $\sum_{i=1}^p \text{tr}[(U \xi_i)(U \xi_i)^T] = \text{tr}[U(\sum_{i=1}^p \xi_i \xi_i^T)U^T] = \text{tr}(UU^T) = \text{tr}(U^T U)$, one gets

$$\begin{aligned} \text{ch}_S(P) \text{tr}(U^T U) &= \text{ch}_S(P) \sum_{i=1}^p \text{tr}[(U \xi_i)(U \xi_i)^T] \\ &\leq \sum_{i=1}^p \eta_i \text{tr}[(U \xi_i)(U \xi_i)^T] \\ &= \text{tr}(PU^TU) \leq \text{ch}_L(P) \sum_{i=1}^p \text{tr}[(U \xi_i)(U \xi_i)^T] \\ &= \text{ch}_L(P) \text{tr}(U^T U). \end{aligned}$$

The lemma follows.

Proof of Theorem 4.2.2.1. In Lemma 4.2.2.1, taking $P = Q_m$

and $U = (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-\frac{1}{2}} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T$, it follows from (4.2.1.3) that

$$\begin{aligned} r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \\ \leq r_0^{-1} m^{-1} \text{ch}_L(Q_m) \text{tr} \left[C (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) \right. \\ \left. \times (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T \right]. \end{aligned} \quad (4.2.2.3)$$

Partitioning $\mathbf{x}_i = (\mathbf{x}_i^{(1)T}, \mathbf{x}_i^{(2)T})^T$, $\mathbf{x}_i^{(1)} = (x_{i1}, \dots, x_{in_i})^T$, $\mathbf{x}_i^{(2)} = (x_{in_i+1}, \dots, x_{iN_i})^T$, $i = 1, \dots, m$, we can write $\Sigma_{11} = \bigoplus_{i=1}^m (I_{n_i} + \lambda_0^{-1} \mathbf{x}_i^{(1)} \mathbf{x}_i^{(1)T})$ and $\Sigma_{21} = \lambda_0^{-1} \bigoplus_{i=1}^m \mathbf{x}_i^{(2)} \mathbf{x}_i^{(1)T}$. Define $w_i = \frac{n_i}{\sum_{j=1}^{n_i} x_{ij}^2} / (\lambda_0 + \frac{n_i}{\sum_{j=1}^{n_i} x_{ij}^2})$. Then

$$\mathcal{C}(\mathbf{X}^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} \mathbf{X}^{(1)}) = \mathbf{1}_{1 \leq i \leq m} [(1 - w_i) f_i \bar{x}_i(u)]$$

where $\bar{x}_i(u) = (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} x_{ij}$ and

$$\begin{aligned} & \text{tr} \left[\mathcal{C}(\mathbf{X}^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} \mathbf{X}^{(1)}) (\mathbf{X}^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} \mathbf{X}^{(1)})^T \mathcal{C}^T \right] \\ &= \sum_{i=1}^m f_i^2 \bar{x}_i^2(u) (1 - w_i)^2 \\ &\leq \sum_{i=1}^m \bar{x}_i^2(u). \end{aligned} \quad (4.2.2.4)$$

$$\begin{aligned} & \mathbf{X}^{(1)T} \Sigma_{11}^{-1} \mathbf{X}^{(1)} \\ &= \sum_{i=1}^m \mathbf{x}_i^{(1)T} (I_{n_i} + \lambda_0^{-1} \mathbf{x}_i^{(1)} \mathbf{x}_i^{(1)T})^{-1} \mathbf{x}_i^{(1)} \\ &= \sum_{i=1}^m \left[\mathbf{x}_i^{(1)T} \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(1)T} \mathbf{x}_i^{(1)} (\mathbf{x}_i^{(1)T} \mathbf{x}_i^{(1)} + \lambda_0)^{-1} \mathbf{x}_i^{(1)T} \mathbf{x}_i^{(1)} \right] \\ &= \sum_{i=1}^m \left(\frac{n_i}{\sum_{j=1}^{n_i} x_{ij}^2} \lambda_0 / \left(\frac{n_i}{\sum_{j=1}^{n_i} x_{ij}^2} + \lambda_0 \right) \right) = \lambda_0 \sum_{i=1}^m w_i. \end{aligned} \quad (4.2.2.5)$$

From (4.2.2.3) - (4.2.2.5), we have

$$r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \leq r_0^{-1} \text{ch}_L(Q_m) \left(m^{-1} \sum_{i=1}^m \bar{x}_{i(u)}^2 \right) \lambda_0^{-1} \left(\sum_{i=1}^m w_i \right)^{-1}. \quad (4.2.2.6)$$

Now, by (4.2.2.2), $m^{-1} \sum_{i=1}^m \bar{x}_{i(u)}^2 = O(1)$ and $\sum_{i=1}^m w_i$ goes to infinity as m goes to infinity. Then by (4.2.2.1), the rhs of (4.2.2.6) goes to zero as $m \rightarrow \infty$ and hence from (4.2.2.6)

$$\overline{\lim}_{m \rightarrow \infty} \left[r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \right] \leq 0. \quad (4.2.2.7)$$

But $r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \geq 0$ for all m and hence

$$\underline{\lim}_{m \rightarrow \infty} \left[r_{Q_m}(\pi_0, \varepsilon_{BF}^*) - r_{Q_m}(\pi_0, \varepsilon_{SB}) \right] \geq 0. \quad (4.2.2.8)$$

Combining (4.2.2.7) and (4.2.2.8), one gets the theorem.

Now, we will consider the sample mean vector. To prove its asymptotic nonoptimality, we need the following conditions in addition to (4.2.2.2).

$$\sup_{i \geq 1} n_i = K \text{ (say)} < \infty, \quad (4.2.2.9)$$

$$N_i \geq n_i + 1 \geq 2 \quad \text{for } i \geq 1, \quad (4.2.2.10)$$

and

$$\underline{\lim}_{m \rightarrow \infty} \text{ch}_S(Q_m) = \omega > 0. \quad (4.2.2.11)$$

Now we state and prove the following theorem.

Theorem 4.2.2.2. Suppose under the prior π_0 , $\bar{Y} \sim N(\bar{X}b_0, r_0^{-1}\Sigma)$. Assume that the conditions given by (4.2.2.2) and (4.2.2.9) - (4.2.2.11) are true. Then, for the random regression coefficients model, the sample mean vector is asymptotically nonoptimal predictor of γ .

Proof of Theorem 4.2.2.2. Let $r_{Q_m}(\pi_0, \bar{Y}_{(s)}) - r_{Q_m}(\pi_0, \varepsilon_{SB}) = d_m$. Since $d_m \geq 0$ for all m , it is enough to prove that $\lim_{m \rightarrow \infty} d_m > 0$.

From (4.2.2.4), using Lemma 4.2.2.1, it follows that

$$\begin{aligned} d_m \geq \text{ch}_S(Q_m)m^{-1} & \left[r_0^{-1} \text{tr}(L - A - C\Sigma_{21}\Sigma_{11}^{-1})\Sigma_{11} \right. \\ & \times (L - A - C\Sigma_{21}\Sigma_{11}^{-1})^T + b_0^2 (LX^{(1)} - AX^{(1)} - CX^{(2)})^T \\ & \left. \times (LX^{(1)} - AX^{(1)} - CX^{(2)}) \right]. \end{aligned} \quad (4.2.2.12)$$

Now,

$$\begin{aligned} & (LX^{(1)} - AX^{(1)} - CX^{(2)})^T (LX^{(1)} - AX^{(1)} - CX^{(2)}) \\ & = \sum_{i=1}^m f_i^2 (\bar{x}_{i(u)} - \bar{x}_{i(s)})^2, \end{aligned} \quad (4.2.2.13)$$

where $\bar{x}_{i(s)} = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$, $i = 1, \dots, m$. Also

$$\begin{aligned} L - A - C\Sigma_{21}\Sigma_{11}^{-1} \\ = \bigoplus_{i=1}^m f_i \left[n_i^{-1} 1_{n_i} - \lambda_0^{-1} (1 - w_i) \bar{x}_{i(u)} x_i^{(1)} \right]^T. \end{aligned}$$

Then

$$\begin{aligned}
& (\underline{L} - \underline{A} - \underline{C}\underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1})\underline{\Sigma}_{11}(\underline{L} - \underline{A} - \underline{C}\underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1})^T \\
& = \sum_{i=1}^m f_i^2 \left[n_i^{-1} + \lambda_0^{-1} (\bar{x}_i(u) - \bar{x}_i(s))^2 - \lambda_0^{-1} (1 - w_i) \bar{x}_i^2(u) \right]
\end{aligned}$$

and hence

$$\begin{aligned}
& \text{tr}(\underline{L} - \underline{A} - \underline{C}\underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1})\underline{\Sigma}_{11}(\underline{L} - \underline{A} - \underline{C}\underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1})^T \\
& = \sum_{i=1}^m f_i^2 \left[n_i^{-1} + \lambda_0^{-1} (\bar{x}_i(u) - \bar{x}_i(s))^2 - \lambda_0^{-1} (1 - w_i) \bar{x}_i^2(u) \right] \\
& = \sum_{i=1}^m f_i^2 \left[n_i^{-1} + \lambda_0^{-1} w_i \bar{x}_i^2(u) - 2\lambda_0^{-1} \bar{x}_i(u) \bar{x}_i(s) + \lambda_0^{-1} \bar{x}_i^2(s) \right] \\
& = \sum_{i=1}^m f_i^2 \left[n_i^{-1} + \lambda_0^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2 + \lambda_0^{-1} (1 - w_i^{-1}) \bar{x}_i^2(s) \right] \\
& = \sum_{i=1}^m f_i^2 \left[n_i^{-1} - \bar{x}_i^2(s) \left(\sum_{j=1}^{n_i} x_{ij}^2 \right)^{-1} + \lambda_0^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2 \right] \\
& = \sum_{i=1}^m f_i^2 \left[\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i(s))^2 / \left(n_i \sum_{j=1}^{n_i} x_{ij}^2 \right) + \lambda_0^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2 \right] \\
& \geq \sum_{i=1}^m f_i^2 \lambda_0^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2. \tag{4.2.2.14}
\end{aligned}$$

From (4.2.2.12) - (4.2.2.14),

$$\begin{aligned}
d_m & \geq \text{ch}_S(\underline{Q}_m) m^{-1} \sum_{i=1}^m f_i^2 \left[(r_0 \lambda_0)^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2 \right. \\
& \quad \left. + b_0^2 (\bar{x}_i(u) - \bar{x}_i(s))^2 \right] \\
& \geq \text{ch}_S(\underline{Q}_m) (K + 1)^{-2} m^{-1} \sum_{i=1}^m \left[(r_0 \lambda_0)^{-1} w_i (\bar{x}_i(u) - w_i^{-1} \bar{x}_i(s))^2 \right. \\
& \quad \left. + b_0^2 (\bar{x}_i(u) - \bar{x}_i(s))^2 \right]. \tag{4.2.2.15}
\end{aligned}$$

Let

$$S_1 = \left\{ i : 1 \leq i \leq m \text{ and } \bar{x}_{i(u)} \leq \frac{1}{2}(1 + w_i^{-1})\bar{x}_{i(s)} \right\}$$

and

$$S_2 = \left\{ i : 1 \leq i \leq m \text{ and } \bar{x}_{i(u)} > \frac{1}{2}(1 + w_i^{-1})\bar{x}_{i(s)} \right\}.$$

Then,

$$\begin{aligned} & \sum_{i=1}^m \left[(r_0 \lambda_0)^{-1} w_i (\bar{x}_{i(u)} - w_i^{-1} \bar{x}_{i(s)})^2 + b_0^2 (\bar{x}_{i(u)} - \bar{x}_{i(s)})^2 \right] \\ & \geq \sum_{i \in S_1} (r_0 \lambda_0)^{-1} w_i \lambda_0^2 \bar{x}_{i(s)}^2 / \left(2 \sum_{j=1}^{n_i} x_{ij}^2 \right)^2 \\ & \quad + \sum_{i \in S_2} b_0^2 \lambda_0^2 \bar{x}_{i(s)}^2 / \left(2 \sum_{j=1}^{n_i} x_{ij}^2 \right)^2 \\ & \geq \lambda_0 r_0^{-1} (4Kp_2^2)^{-1} (\lambda_0 + Kp_2^2)^{-1} \sum_{i \in S_1} \bar{x}_{i(s)}^2 \\ & \quad + b_0^2 \lambda_0^2 (4K^2 p_2^4)^{-1} \sum_{i \in S_2} \bar{x}_{i(s)}^2, \quad \text{by (4.2.2.2), (4.2.2.9)} \\ & \quad \text{and (4.2.2.10)} \\ & \geq d \sum_{i=1}^m \bar{x}_{i(s)}^2 \geq dp_1^2 m, \quad (4.2.2.16) \end{aligned}$$

by (4.2.2.2) where

$$d = \min \left(\lambda_0 r_0^{-1} (4Kp_2^2)^{-1} (\lambda_0 + Kp_2^2)^{-1}, b_0^2 \lambda_0^2 (4K^2 p_2^4)^{-1} \right).$$

From (4.2.2.15) and (4.2.2.16) we have

$$d_m \geq ch_S(Q_m) m^{-1} (K + 1)^{-2} dp_1^2 m$$

and by (4.2.2.11)

$$\lim_{m \rightarrow \infty} d_m \geq \omega(K+1)^{-2} dp_1^2 > 0 \quad \text{provided } b_0 \neq 0.$$

This completes the proof of Theorem 4.2.2.2.

Remark 4.2.2.2. Note that for average squared error loss $Q_m = I_m$, and (4.2.2.1) and (4.2.2.11) are satisfied. Then from the preceding two theorems it follows that while the sample mean vector is asymptotically nonoptimal, the HB predictor is asymptotically optimal.

4.2.3 Nested Error Regression Model

In this subsection we will examine the asymptotic behavior of the risk differences given in (4.2.1.3) and (4.2.1.4) under the nested error regression model. The corresponding linear model is given by the equation (2.2.3). Here $\Psi = I_{N_T}$, $\lambda_0 = \lambda_0$, $D(\lambda_0) = \lambda_0^{-1} I_m$, $Z^{(1)} = \bigoplus_{i=1}^m 1_{n_i}$, $Z^{(2)} = \bigoplus_{i=1}^m 1_{N_i - n_i}$ and $\Sigma = \bigoplus_{i=1}^m (I_{N_i} + \lambda_0^{-1} J_{N_i})$. Also $X^{(1)} = \text{col}_{1 \leq i \leq m} \text{col}_{1 \leq j \leq n_i} (x_{ij}^T)$ and $X^{(2)} = \text{col}_{1 \leq i \leq m} \text{col}_{n_i+1 \leq j \leq N_i} (x_{ij}^T)$.

We will prove here that under suitable conditions, given below, the risk difference between the HB and the subjective Bayes predictor goes to zero as $m \rightarrow \infty$ whereas the positive risk difference between the sample mean vector and the subjective Bayes predictor remains bounded away from zero. These two results will be provided in the form of theorems. The first theorem in this subsection proves the A.O. property of the HB predictor whereas the second

theorem proves a negative result about the sample mean vector which is design unbiased for the finite population mean vector under simple random sampling.

To prove Theorem 4.2.3.1, we need to assume that

$\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s)$ is p.d. Moreover,

$$\lim_{m \rightarrow \infty} \text{ch}_L \left[\left(\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) / m \right)^{-1} \right] < \infty, \quad (4.2.3.1)$$

$$\sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) = o(m), \quad (4.2.3.2)$$

and

$$\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) = o(m) \quad (4.2.3.3)$$

where $\bar{x}_i(s) = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$, $\bar{x}_i(u) = (N_i - n_i)^{-1} \sum_{j=n_i+1}^{N_i} x_{ij}$,
 $i = 1, \dots, m$.

Remark 4.2.3.1. Note that if for some p.d. matrix Q ,

$\lim_{m \rightarrow \infty} \left(\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) / m \right) = Q$, then conditions (4.2.3.1) and (4.2.3.2) hold.

Also to prove Theorem 4.2.3.1 below, we need an inequality involving the eigen values of matrices. To this end, we state and prove the following lemma.

Lemma 4.2.3.1. For a symmetric n.n.d. matrix $P(p \times p)$ and a symmetric matrix $T(p \times p)$

- (i) $\text{ch}_L(\underline{T}) \leq \text{ch}_L(\underline{P} + \underline{T})$
(ii) $\text{ch}_S(\underline{T}) \leq \text{ch}_S(\underline{P} + \underline{T})$.

Proof of Lemma 4.2.3.1. From p. 62 of Rao (1973), we have

$$\text{ch}_L(\underline{T}) = \sup_{\underline{a} \neq 0} \frac{\underline{a}^T \underline{T} \underline{a}}{\underline{a}^T \underline{a}} \leq \sup_{\underline{a} \neq 0} \frac{\underline{a}^T (\underline{P} + \underline{T}) \underline{a}}{\underline{a}^T \underline{a}} = \text{ch}_L(\underline{P} + \underline{T}).$$

Similarly (ii) can be proved.

Theorem 4.2.3.1. Suppose the prior π_0 is given by $\underline{Y} \sim N(\underline{X}\underline{b}_0, r_0^{-1}\underline{\Sigma})$, and the conditions given by (4.2.2.1), (4.2.2.9) and (4.2.3.1) - (4.2.3.3) are true. Then $\underline{e}_{BF}^*(\underline{Y}^{(1)})$, the HB predictor of $\underline{\gamma}$ under the nested error regression model, is asymptotically optimal under the loss (4.2.2).

Proof of Theorem 4.2.3.1. Let $a_m = r_{Q_m}(\pi_0, \underline{e}_{BF}^*) - r_{Q_m}(\pi_0, \underline{e}_{SB})$. Note that $a_m \geq 0$ for all m . To prove the theorem, it is enough to show $\overline{\lim}_{m \rightarrow \infty} a_m = 0$.

Now, from (4.2.1.3), taking $\underline{P} = Q_m$ and $\underline{U} = (\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)})^{-\frac{1}{2}} (\underline{X}^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)})^T \underline{C}^T$ in Lemma 4.2.2.1, it follows that

$$\begin{aligned} a_m \leq r_0^{-1} m^{-1} \text{ch}_L(Q_m) \text{tr} \left[\underline{C} (\underline{X}^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)}) \right. \\ \left. \times (\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)})^{-1} (\underline{X}^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)})^T \underline{C}^T \right]. \end{aligned} \quad (4.2.3.4)$$

Now taking $P = (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1}$ and $U^T = (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T$ in the Lemma 4.2.2.1, we have from (4.2.3.4) that

$$a_m \leq r_0^{-1} m^{-1} \text{ch}_L(Q_m) \text{ch}_L(X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} \\ \times \text{tr} \left[C(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T \right]. \quad (4.2.3.5)$$

Here $\Sigma_{21} \Sigma_{11}^{-1} = \bigoplus_{i=1}^m \left[(n_i + \lambda_0)^{-1} J_{N_i - n_i, n_i} \right]$, $\Sigma_{21} \Sigma_{11}^{-1} X^{(1)} =$
 $\bigoplus_{i=1}^m \left(n_i (n_i + \lambda_0)^{-1} 1_{N_i - n_i} \bar{x}_i^T(s) \right)$ and $X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} =$
 $\bigoplus_{i=1}^m \left[\bigoplus_{j=1}^{n_i} (x_{ij}^T - n_i (n_i + \lambda_0)^{-1} 1_{N_i - n_i} \bar{x}_i^T(s)) \right]$. From
 these we have

$$\text{tr} \left[C(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T C^T \right] \\ = \sum_{i=1}^m f_i^2 (\bar{x}_i(u) - n_i (n_i + \lambda_0)^{-1} \bar{x}_i(s))^T (\bar{x}_i(u) - n_i (n_i + \lambda_0)^{-1} \bar{x}_i(s)) \\ \leq \sum_{i=1}^m (\bar{x}_i(u) - n_i (n_i + \lambda_0)^{-1} \bar{x}_i(s))^T (\bar{x}_i(u) - n_i (n_i + \lambda_0)^{-1} \bar{x}_i(s)) \\ \leq 2 \left[\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) + \sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) \right], \quad (4.2.3.6)$$

since for two vectors a_1 ($p \times 1$), a_2 ($p \times 1$) $(a_1 - a_2)^T (a_1 - a_2) \leq 2(a_1^T a_1 + a_2^T a_2)$ and $n_i (n_i + \lambda_0)^{-1} < 1$ for $i = 1, \dots, m$.

From (2.4.2) we have

$$X^{(1)T} \Sigma_{11}^{-1} X^{(1)} = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i(s)) (x_{ij} - \bar{x}_i(s))^T \\ + \lambda_0 \sum_{i=1}^m n_i (n_i + \lambda_0)^{-1} \bar{x}_i(s) \bar{x}_i^T(s).$$

From this and (4.2.2.9) we get that $\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)}$
 $- \lambda_0(K + \lambda_0)^{-1} \sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s)$ is n.n.d. and hence from Lemma
 4.2.3.1, taking $\underline{P} = \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} - \lambda_0(K + \lambda_0)^{-1} \sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s)$
 and $\underline{T} = \lambda_0(K + \lambda_0)^{-1} \sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s)$, we have

$$\text{ch}_S \left(\lambda_0(K + \lambda_0)^{-1} \sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) \right) \leq \text{ch}_S \left(\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right)$$

i.e.,

$$\text{ch}_L \left[\left(\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right)^{-1} \right] \leq (1 + K/\lambda_0) \text{ch}_L \left[\left(\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) \right)^{-1} \right].$$

(4.2.3.7)

From (4.2.3.5) - (4.2.3.7), we have

$$\begin{aligned} a_m &\leq 2r_0^{-1} m^{-1} \text{ch}_L(\underline{Q}_m) (1 + K/\lambda_0) \text{ch}_L \left(\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) \right)^{-1} \\ &\quad \times \left[\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) + \sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) \right] \\ &= 2r_0^{-1} (1 + K/\lambda_0) \text{ch}_L(\underline{Q}_m) \text{ch}_L \left(\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) / m \right)^{-1} \\ &\quad \times \left[\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) / m^2 + \sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) / m^2 \right]. \end{aligned} \quad (4.2.3.8)$$

The rhs of (4.2.3.8) goes to zero as m goes to infinity, by conditions (4.2.2.1) and (4.2.3.1) - (4.2.3.3). Hence from (4.2.3.8)

$$\overline{\lim}_{m \rightarrow \infty} a_m \leq 0.$$

But $a_m \geq 0$ for all m . Consequently $\overline{\lim}_{m \rightarrow \infty} a_m = 0$ and the proof of Theorem 4.2.3.1 is complete.

Remark 4.2.3.2. Note that Theorem 4.2.3.1 remains true if we assume $\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) = o(m^2)$ and $\sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) = o(m^2)$ or if we replace (4.2.3.2) by $\sum_{i=1}^m \bar{x}_i^T(s) \bar{x}_i(s) = O(m^{1+\epsilon})$ and (4.2.3.3) by $\sum_{i=1}^m \bar{x}_i^T(u) \bar{x}_i(u) = O(m^{1+\epsilon})$ for any $\epsilon < 1$. In a special case of the nested error regression model, to be considered in Section 4.4, $\bar{x}_{ij} = \bar{x}_i$, $j = 1, \dots, N_i$, $i = 1, \dots, m$; in this case conditions (4.2.3.2) and (4.2.3.3) are identical. Also in this case $\sum_{i=1}^m \bar{x}_i(s) \bar{x}_i^T(s) = \sum_{i=1}^m \bar{x}_i \bar{x}_i^T$ is p.d. since it is assumed that $\text{rank}(\bar{X}^{(1)}) = p$.

Now we will briefly consider the infinite population set up for this model. An alternative predictor for γ in the nested error regression model, as we have seen in Section 2.4, is the predictor of $\bar{X}b + \gamma = \mu$ (say), the conditional mean vector given the values of the covariates and the realized value of the random vector γ of the stratum effect, where $\bar{X} = \underset{1 \leq i \leq m}{\text{col}} \left(\bar{x}_i^T(p) \right)$ and $\bar{x}_i(p) = N_i^{-1} \sum_{j=1}^{N_i} \bar{x}_{ij}$. From Section 3.2 one can show that the HB predictor $\hat{\mu}_{HB}$ (say) of μ under the loss (4.2.2) is given by

$$\begin{aligned} \hat{\mu}_{HB} = & \left[\bar{X} - \sum_{1 \leq i \leq m}^{col} \left(n_i (n_i + \lambda_0)^{-1} \bar{x}_{i(s)}^T \right) \right] \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\ & \times X^{(1)T} \Sigma_{11}^{-1} Y^{(1)} + \sum_{1 \leq i \leq m}^{col} \left(n_i (n_i + \lambda_0)^{-1} \bar{y}_{i(s)} \right). \quad (4.2.3.9) \end{aligned}$$

Now assume that the prior π_0 specifies that $\begin{pmatrix} Y^{(1)} \\ y \end{pmatrix}$ is MVN with $E_{\pi_0}(Y^{(1)}) = X^{(1)} b_0$, $E_{\pi_0}(y) = 0$, $V_{\pi_0}(Y^{(1)}) = r_0^{-1} \Sigma_{11}$, $V_{\pi_0}(y) = (r_0 \lambda_0)^{-1} I_m$ and $Cov_{\pi_0}(Y^{(1)}, y) = (r_0 \lambda_0)^{-1} Z^{(1)}$. Then the subjective Bayes predictor $\hat{\mu}_{SB}$ (say) of μ under the loss (4.2.2) is given by

$$\begin{aligned} \hat{\mu}_{SB} = & \left[\bar{X} - \sum_{1 \leq i \leq m}^{col} \left(n_i (n_i + \lambda_0)^{-1} \bar{x}_{i(s)}^T \right) \right] b_0 \\ & + \sum_{1 \leq i \leq m}^{col} \left(n_i (n_i + \lambda_0)^{-1} \bar{y}_{i(s)} \right). \quad (4.2.3.10) \end{aligned}$$

It can be shown that

$$\begin{aligned} r_{Q_m}(\pi_0, \hat{\mu}_{HB}) - r_{Q_m}(\pi_0, \hat{\mu}_{SB}) \\ = (mr_0)^{-1} \text{tr} \left[Q_m \left\{ \bar{X} - \sum_{1 \leq i \leq m}^{col} \left(n_i (\lambda_0 + n_i)^{-1} \bar{x}_{i(s)}^T \right) \right\} \right. \\ \left. \times \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \left\{ \bar{X} - \sum_{1 \leq i \leq m}^{col} \left(n_i (\lambda_0 + n_i)^{-1} \bar{x}_{i(s)}^T \right) \right\}^T \right] \end{aligned}$$

and that under the same conditions of Theorem 4.2.3.1 this risk difference tends to 0 as $m \rightarrow \infty$. Thus, $\hat{\mu}_{HB}$ also possesses the A.O. property. Now in the following theorem

we will establish that the sample mean vector $\bar{Y}_{(s)}$ is asymptotically nonoptimal under certain conditions.

Theorem 4.2.3.2. Under the prior π_0 given by $Y \sim N(Xb_0, r_0^{-1}\Sigma)$ and the conditions (4.2.2.9) - (4.2.2.11), for the nested error regression model the sample mean vector $\bar{Y}_{(s)}$ is asymptotically nonoptimal.

Proof of Theorem 4.2.3.2. As in Theorem 4.2.2.2, it is enough to show that for $d_m = r_{Q_m}(\pi_0, \bar{Y}_{(s)}) - r_{Q_m}(\pi_0, e_{SB})$

$$\lim_{m \rightarrow \infty} d_m > 0.$$

From (4.2.1.4) and Lemma 4.2.2.1 we have

$$d_m \geq r_0^{-1} m^{-1} \text{ch}_S(Q_m) \text{tr} \left[(L - A - C\Sigma_{21}\Sigma_{11}^{-1})\Sigma_{11}(L - A - C\Sigma_{21}\Sigma_{11}^{-1})^T \right]. \quad (4.2.3.11)$$

In this case, $L - A - C\Sigma_{21}\Sigma_{11}^{-1} = \lambda_0 \bigoplus_{i=1}^m f_i(\lambda_0 + n_i)^{-1} n_i^{-1} 1_{n_i}^T$ and hence $(L - A - C\Sigma_{21}\Sigma_{11}^{-1})\Sigma_{11}(L - A - C\Sigma_{21}\Sigma_{11}^{-1})^T = \lambda_0 \text{Diag}(f_1^2 n_1^{-1}(\lambda_0 + n_1)^{-1}, \dots, f_m^2 n_m^{-1}(\lambda_0 + n_m)^{-1})$. Then from (4.2.3.11) we have

$$\begin{aligned} d_m &\geq r_0^{-1} m^{-1} \text{ch}_S(Q_m) \lambda_0 \sum_{i=1}^m f_i^2 n_i^{-1} (\lambda_0 + n_i)^{-1} \\ &\geq r_0^{-1} \lambda_0 (\lambda_0 + K)^{-1} K^{-1} (K + 1)^{-2} \text{ch}_S(Q_m) \end{aligned}$$

by (4.2.2.9) and (4.2.2.10). Hence, by (4.2.2.11)

$$\lim_{m \rightarrow \infty} d_m \geq r_0^{-1} \lambda_0 (\lambda_0 + K)^{-1} K^{-1} (K + 1)^{-2} \omega > 0$$

and the theorem follows.

4.3 Asymptotic Optimality with Known First Stage Variance Component: Fay-Herriot Model

Consider the following hierarchical model.

I. Conditional on $\theta_i = \theta_i$, $B = b$ and $\Delta = \delta$,

$$\underline{Y}_i = (Y_{i1}, \dots, Y_{iN_i})^T \sim N(\theta_i \underline{1}_{N_i}, V_i \underline{I}_{N_i}),$$

$i = 1, \dots, m$ independently where $V_i (>0)$ is known sampling variance for i^{th} stratum;

II. conditional on $B = b = (b_1, \dots, b_p)^T$, and $\Delta = \delta$,

$$\underline{\Theta} \sim N(\underline{X}_* b, \delta \underline{I}_m),$$

where $\underline{\Theta} = (\theta_1, \dots, \theta_m)^T$, $\underline{\theta} = (\theta_1, \dots, \theta_m)^T$ and $\underline{X}_* = (\underline{x}_1, \dots, \underline{x}_m)^T$, each $\underline{x}_i$ being a p -component column vector. It is assumed that $m > p$ and $\text{rank}(\underline{X}_*) = p$.

III. B and Δ are marginally independent with $B \sim \text{uniform}(R^p)$, and Δ having a (proper or improper) distribution $h(\delta)$ on $(0, \infty)$. Explicit form of $h(\delta)$ will be provided soon.

The model given above by I and II is known as Fay-Herriot model and was referred to in Section 2.5. Here we would like to show that the HB predictor of $\underline{\gamma}$, the finite population mean vector, is asymptotically optimal under the loss (4.2.2). We will accomplish this by showing that the HB predictor of $\underline{\theta}$ is asymptotically optimal. Assume that we have the sample mean vector $\bar{Y}_{(s)} = (\bar{Y}_{1(s)}, \dots, \bar{Y}_{m(s)})^T$ where $\bar{Y}_{i(s)}$ is based on n_i observations, $i = 1, \dots, m$. Assume n_i are so chosen that $V_i/n_i = V$ for all $i = 1, \dots, m$. In that case, from I-II we have

I'. conditional on $\Theta = \underline{\theta}$, $B = \underline{b}$ and $\Delta = \delta$,

$$\bar{Y}_{(s)} \sim N(\underline{\theta}, V\underline{I}_m),$$

II'. conditional on $B = \underline{b}$ and $\Delta = \delta$,

$$\Theta \sim N(\underline{X}_* \underline{b}, \delta \underline{I}_m);$$

and we replace III by

III'. B and Δ are marginally independent with $B \sim \text{uniform}(R^p)$, and Δ having a Type II (proper or improper) beta pdf given by

$$h(\delta) \propto \delta^{\alpha-1} (V + \delta)^{-(\alpha+\beta)} I_{[0 < \delta < \infty]}, \quad (4.3.1)$$

with $\alpha > 0$ and β (real). In order to ensure a proper posterior distribution for Δ , we shall impose some restriction on β later.

Some authors considered models which are either special cases or generalizations in some sense of the model given by I'-III' with $\underline{Y}$ replacing $\bar{Y}_{(s)}$. For example, in Morris (1981), the case $p = 1$ and $\underline{X}_* = \underline{1}_m$ was considered. He used the notation μ for b_1 and obtained HB estimators of $\underline{\theta}$ when $\alpha = 1$ and $\beta = -1$. In Morris (1983), stages I' and II' of the model are considered in their full generality, but $\underline{b}$ is assumed known, and an EB rather than an HB procedure is used in the sense that no distribution on Δ is assigned, and it is estimated on the basis of the marginal distribution of $\underline{Y}$. Ghosh (1989) considered all the three stages of the model with $\alpha = 1$ and $\beta = -1$.

Strawderman (1971) considered the model with known $\underline{b}$ and $\alpha = 1$, $\beta > 0$, while Faith (1978) considered the case of known $\underline{b}$, but assigned a general class of priors to Δ including but not limited to the Type-II beta density as given in (4.3.1) with $\alpha \geq 1$ and $\beta > 0$.

Throughout we assume without mention that m is so large that $\frac{1}{2}(m - p) + \beta > 0$. Algebraic manipulations lead to the following facts from I'-III':

(i) conditional on $\bar{Y}_{(s)} = \bar{y}_{(s)}$ and $\Delta = \delta$,

$$\underline{\Theta} \sim N(\underline{H}\bar{y}_{(s)}, V\underline{H}), \quad (4.3.2)$$

where $\underline{H} = V^{-1}(V^{-1} + \delta^{-1})^{-1}(\underline{I}_m + V\delta^{-1}\underline{P}_{\underline{X}_*})$, $\underline{P}_{\underline{X}_*} = \underline{X}_*(\underline{X}_*^T \underline{X}_*)^{-1} \underline{X}_*^T$;

(ii) conditional on $\bar{Y}_{(s)} = \bar{y}_{(s)}$, Δ has pdf

$$\begin{aligned} f(\delta | \bar{y}_{(s)}) &\propto \exp \left[-\frac{1}{2}(V + \delta)^{-1} \left\{ \bar{y}_{(s)}^T (\underline{I}_m - \underline{P}_{\underline{X}_*}) \bar{y}_{(s)} \right\} \right] \\ &\times (V + \delta)^{-\frac{1}{2}(m-p)} \delta^{\alpha-1} (V + \delta)^{-(\alpha+\beta)}, \quad 0 < \delta < \infty. \end{aligned} \quad (4.3.3)$$

Let $U = V/(V + \Delta)$ and $u = V/(V + \delta)$. Then it follows from (4.3.2) that

$$E(\underline{\Theta} | u, \bar{y}_{(s)}) = (1 - u)\bar{y}_{(s)} + u\underline{P}_{\underline{X}_*}\bar{y}_{(s)}. \quad (4.3.4)$$

Also, writing $s_m = \bar{y}_{(s)}^T (\underline{I}_m - \underline{P}_{\underline{X}_*}) \bar{y}_{(s)} / V$, it follows from

(4.3.3) that the conditional pdf of U given $\bar{Y}_{(s)} = \bar{y}_{(s)}$ is

$$f_{\alpha, \beta}(u | \bar{y}_{(s)}) \propto \exp\left(-\frac{1}{2}uS_m\right) u^{\frac{1}{2}(m-p)+\beta-1} (1-u)^{\alpha-1} I_{[0 < u < 1]}. \quad (4.3.5)$$

Accordingly, under the quadratic loss $L(\theta, \underline{a})$ with L given in (4.2.2), the HB estimator of θ is given by

$$\begin{aligned} \varepsilon_{BI}(\bar{Y}_{(s)}) &= E(\Theta | \bar{Y}_{(s)}) \\ &= \left(1 - E_{\alpha, \beta}(U | \bar{Y}_{(s)})\right) \bar{Y}_{(s)} + E_{\alpha, \beta}(U | \bar{Y}_{(s)}) P_{X_*} \bar{Y}_{(s)}, \end{aligned} \quad (4.3.6)$$

where

$$\begin{aligned} E_{\alpha, \beta}(U | \bar{Y}_{(s)}) &= \int_0^1 u f_{\alpha, \beta}(u | \bar{Y}_{(s)}) du \\ &= \int_0^1 u^{\frac{1}{2}(m-p)+\beta} (1-u)^{\alpha-1} \exp\left(-\frac{1}{2}uS_m\right) du \\ &\quad \div \int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} (1-u)^{\alpha-1} \exp\left(-\frac{1}{2}uS_m\right) du \end{aligned} \quad (4.3.7)$$

and

$$S_m = \bar{Y}_{(s)}^T (I_m - P_{X_*}) \bar{Y}_{(s)} / V.$$

In the above $E_{\alpha, \beta}$ denotes the expectation w.r.t. the pdf $f_{\alpha, \beta}$ given in (4.3.5).

We shall now evaluate the performance of the HB estimator ε_{BI} of θ for large m under the loss (4.2.2) and the $N(X_* b_0, \delta_0 I_m)$ prior for Θ . The subjective Bayes

estimator of θ under this prior, say π_0 and the loss (4.2.2) is given by

$$e_{SB}^*(\bar{Y}_{(s)}) = (1 - u_0)\bar{Y}_{(s)} + u_0 X^* b_0, \quad (4.3.8)$$

where $u_0 = V/(V + \delta_0)$.

In this set up, for an estimator $\hat{\theta}$ based on $\bar{Y}_{(s)}$ of θ , we denote its Bayes risk under the prior π_0 and the loss L in (4.2.2) by

$$\begin{aligned} r_{Q_m}(\pi_0, \hat{\theta}) \\ = E_{\pi_0} \left[E_{\hat{\theta}} \left\{ \left(\hat{\theta}(\bar{Y}_{(s)}) - \theta \right)^T m^{-1} Q_m \left(\hat{\theta}(\bar{Y}_{(s)}) - \theta \right) \right\} \right] \end{aligned} \quad (4.3.9)$$

where the expectation $E_{\hat{\theta}}$ inside the square bracket is w.r.t. the conditional distribution $\bar{Y}_{(s)} \sim N(\theta, V I_m)$ and the outer expectation E_{π_0} is w.r.t. the prior distribution π_0 . As in (4.2.1.1), we have

$$\begin{aligned} r_{Q_m}(\pi_0, \hat{\theta}) - r_{Q_m}(\pi_0, e_{SB}^*) \\ = m^{-1} E_{\pi_0} \left[E_{\hat{\theta}} \left\{ \left(\hat{\theta}(\bar{Y}_{(s)}) - e_{SB}^*(\bar{Y}_{(s)}) \right)^T Q_m \left(\hat{\theta}(\bar{Y}_{(s)}) - e_{SB}^*(\bar{Y}_{(s)}) \right) \right\} \right] \\ = m^{-1} E^* \left[\left(\hat{\theta}(\bar{Y}_{(s)}) - e_{SB}^*(\bar{Y}_{(s)}) \right)^T Q_m \left(\hat{\theta}(\bar{Y}_{(s)}) - e_{SB}^*(\bar{Y}_{(s)}) \right) \right] \end{aligned} \quad (4.3.10)$$

where the expectation E^* is w.r.t. the marginal

distribution $m_{\pi_0}(\bar{Y}_{(s)})$ of $\bar{Y}_{(s)} \sim N(X_*b_0, (V + \delta_0)I_m)$ obtained by averaging the (conditional) distribution of $\bar{Y}_{(s)}$ given in I' w.r.t. the prior π_0 . We first show that $r_{Q_m}(\pi_0, \varepsilon_{BI}) - r_{Q_m}(\pi_0, \varepsilon_{SB}^*) \rightarrow 0$ as $m \rightarrow \infty$. To achieve this, we need to prove a series of lemmas. Denote the expectation $E_{\alpha, \beta}(U|\bar{Y}_{(s)})$ given in (4.3.7) by $T_m(\bar{Y}_{(s)}; \alpha, \beta)$, that is

$$\begin{aligned} T_m(\bar{Y}_{(s)}; \alpha, \beta) &= \int_0^1 u^{\frac{1}{2}(m-p)+\beta} (1-u)^{\alpha-1} \exp\left(-\frac{1}{2}uS_m\right) du \\ &\div \int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} (1-u)^{\alpha-1} \exp\left(-\frac{1}{2}uS_m\right) du. \end{aligned} \quad (4.3.11)$$

We subscripted it by m to show its dependence on m .

Note that $T_m(\bar{Y}_{(s)}; \alpha, \beta)$ is a statistic since α and β are known prior parameters as appeared in III' . The first lemma of this section shows that for $\alpha = 1$, $T_m(\bar{Y}_{(s)}; 1, \beta) \xrightarrow{a.s.} u_0$ when θ has the prior π_0 , i.e., marginally $\bar{Y}_{(s)} \sim N(X_*b_0, (V + \delta_0)I_m)$.

Lemma 4.3.1. Suppose $\bar{Y}_{(s)} \sim N(X_*b_0, (V + \delta_0)I_m)$. Then $T_m(\bar{Y}_{(s)}; 1, \beta)$ converges a.s. (as $m \rightarrow \infty$) to u_0 .

Proof of Lemma 4.3.1. For $\alpha = 1$, from (4.3.11) we have

$$\begin{aligned}
& T_m(\bar{Y}_{(s)}; 1, \beta) \\
&= \int_0^1 u^{\frac{1}{2}(m-p)+\beta} \exp\left(-\frac{1}{2}uS_m\right) du \Big/ \int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} \exp\left(-\frac{1}{2}uS_m\right) du.
\end{aligned}
\tag{4.3.12}$$

First, integrating by parts the numerator in (4.3.12), it follows from (4.3.12) that

$$\begin{aligned}
T_m(\bar{Y}_{(s)}; 1, \beta) &= (m - p + 2\beta)/S_m - \left(\frac{1}{2}S_m \exp\left(\frac{1}{2}S_m\right)\right)^{-1} \\
&\quad \times \left[\int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} \exp\left(-\frac{1}{2}uS_m\right) du \right]^{-1}.
\end{aligned}
\tag{4.3.13}$$

Assume first that $m - p$ is an even integer and β is an integer. Using the symbol $(d)_r = d(d - 1) \dots (d - r + 1)$ for $d \geq r$, successive integration by parts leads to

$$\begin{aligned}
& \int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} \exp\left(-\frac{1}{2}uS_m\right) du \\
&= \left(\frac{1}{2}S_m\right)^{-\left(\frac{1}{2}(m-p)+\beta\right)} \left(\frac{1}{2}(m - p) + \beta - 1\right)! \\
&\quad - \sum_{r=0}^{\frac{1}{2}(m-p)+\beta-1} \left(\frac{1}{2}(m - p) + \beta - 1\right)_r \\
&\quad \times \left(\frac{1}{2}S_m\right)^{-r-1} \exp\left(-\frac{1}{2}S_m\right).
\end{aligned}
\tag{4.3.14}$$

Accordingly,

$$\begin{aligned}
& \left(\frac{1}{2}S_m\right)\exp\left(\frac{1}{2}S_m\right) \times (\text{lhs of (4.3.14)}) \\
&= \left(\frac{1}{2}S_m\right)^{-\left(\frac{1}{2}(m-p)+\beta-1\right)} \left(\frac{1}{2}(m-p) + \beta - 1\right)! \exp\left(\frac{1}{2}S_m\right) \\
&\quad - \sum_{r=0}^{\frac{1}{2}(m-p)+\beta-1} \left(\frac{1}{2}(m-p) + \beta - 1\right)_r \left(\frac{1}{2}S_m\right)^{-r}.
\end{aligned} \tag{4.3.15}$$

Next writing $S_m = (\bar{Y}_{(S)} - X_*b_0)^T(I_m - P_{X_*})(\bar{Y}_{(S)} - X_*b_0)/V$, and noting the idempotency of $I_m - P_{X_*}$, it follows that $S_m \sim u_0^{-1}\chi_{m-p}^2$. Now using Markov's inequality, for every $\epsilon > 0$,

$$\begin{aligned}
& P\left(\left|S_m/(u_0^{-1}(m-p)) - 1\right| > \epsilon\right) \\
&\leq \epsilon^{-4} E\left[\left\{S_m/(u_0^{-1}(m-p)) - 1\right\}^4\right] \\
&= O(m^{-2}).
\end{aligned} \tag{4.3.16}$$

Application of Borel-Cantelli Lemma now leads to

$S_m/(u_0^{-1}(m-p)) \xrightarrow{a.s.} 1$ as $m \rightarrow \infty$. Hence,

$$\begin{aligned}
& \sum_{r=0}^{\frac{1}{2}(m-p)+\beta-1} \left(\frac{1}{2}(m-p) + \beta - 1\right)_r \left(\frac{1}{2}S_m\right)^{-r} \\
&\leq \sum_{r=0}^{\frac{1}{2}(m-p)+\beta-1} \left[\left(\frac{1}{2}(m-p) + \beta - 1\right)^r / \left(\frac{1}{2}(m-p)\right)^r \right] (S_m/(m-p))^{-r} \\
&\leq \sum_{r=0}^{\infty} \left[\left(\frac{1}{2}(m-p) + \beta - 1\right)^r / \left(\frac{1}{2}(m-p)\right)^r \right] (S_m/(m-p))^{-r}.
\end{aligned} \tag{4.3.17}$$

Choose m_0 so large that $2(\beta - 1)/(m_0 - p) < g_0$ where $1 + g_0 < u_0^{-1}$. Hence, for $m \geq m_0$,

$$\begin{aligned} \text{rhs of (4.3.17)} &\leq \sum_{r=0}^{\infty} (1 + g_0)^r (S_m/(m - p))^{-r} \\ &\stackrel{\text{a.s.}}{\leq} \sum_{r=0}^{\infty} (u_0(1 + g_0))^r < \infty. \end{aligned} \quad (4.3.18)$$

Also, let $\ell(m) = \left(\frac{1}{2}S_m\right)^{-\left(\frac{1}{2}(m-p)+\beta-1\right)} \left(\frac{1}{2}(m-p) + \beta - 1\right)! \times \exp\left(\frac{1}{2}S_m\right)$. Hence,

$$\begin{aligned} \log \ell(m) &= \frac{1}{2}S_m + \sum_{j=2}^{\frac{1}{2}(m-p)+\beta-1} \log j - \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(\frac{1}{2}S_m\right) \\ &\geq \frac{1}{2}S_m + \int_1^{\frac{1}{2}(m-p)+\beta-1} \log x \, dx - \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(\frac{1}{2}S_m\right) \\ &= \frac{1}{2}S_m + \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(\frac{1}{2}(m-p) + \beta - 1\right) \\ &\quad - \left(\frac{1}{2}(m-p) + \beta - 1\right) + 1 - \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(\frac{1}{2}S_m\right) \\ &= \frac{1}{2}S_m + \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(1 + \frac{2(\beta - 1)}{m - p}\right) \\ &\quad - \frac{1}{2}(m-p) - (\beta - 2) - \left(\frac{1}{2}(m-p) + \beta - 1\right) \log\left(\frac{\frac{1}{2}S_m}{\frac{1}{2}(m-p)}\right). \end{aligned} \quad (4.3.19)$$

Since $S_m/(m - p) \stackrel{\text{a.s.}}{\rightarrow} u_0^{-1}$, it follows from (4.3.19) that

$$\lim_{m \rightarrow \infty} m^{-1} \log \ell(m) \geq \frac{1}{2}(u_0^{-1} - 1 - \log u_0^{-1}) > 0 \text{ a.s.} \quad (4.3.20)$$

Hence, from (4.3.15) and (4.3.17) - (4.3.20), one gets that lhs of (4.3.15) $\rightarrow \infty$ a.s. as $m \rightarrow \infty$. Once again, recalling that $S_m/(m-p) \xrightarrow{a.s.} u_0^{-1}$ as $m \rightarrow \infty$, it follows from (4.3.13) and (4.3.14) that $T_m(\bar{Y}_{(s)}; 1, \beta) \xrightarrow{a.s.} u_0$ as $m \rightarrow \infty$.

If $m-p$ is odd and β is an integer, use the inequality

$$\begin{aligned} & \int_0^1 u^{\frac{1}{2}(m-p)+\beta-1} \exp\left(-\frac{1}{2}uS_m\right) du \\ & \geq \int_0^1 u^{\frac{1}{2}(m-p+1)+\beta-1} \exp\left(-\frac{1}{2}uS_m\right) du, \end{aligned} \quad (4.3.21)$$

and proceed similarly as before to conclude that

$$\begin{aligned} T_m(\bar{Y}_{(s)}; 1, \beta) & \xrightarrow{a.s.} u_0 \text{ as } m \rightarrow \infty. \text{ Thus, for integer } \beta, \\ T_m(\bar{Y}_{(s)}; 1, \beta) & \xrightarrow{a.s.} u_0 \text{ as } m \rightarrow \infty. \end{aligned}$$

For noninteger β , we make the following observation.

Denote $f_{1,\beta}(u|\bar{Y}_{(s)})$ by $f_\beta(u)$ for brevity. Hence for $\beta < \beta'$, $f_{\beta'}(u)/f_\beta(u) \propto u^{\beta'-\beta} \uparrow$ in u . Using this monotone likelihood ratio (MLR) property, it follows from Lemma 2, part (i) of Lehmann (1986, p. 85) that if $[\beta]$ denotes the integer not exceeding β

$$\begin{aligned} E_{1, [\beta]}(U|\bar{Y}_{(s)}) & \leq E_{1, \beta}(U|\bar{Y}_{(s)}) \leq E_{1, [\beta]+1}(U|\bar{Y}_{(s)}) \\ \text{i.e., } T_m(\bar{Y}_{(s)}; 1, [\beta]) & \leq T_m(\bar{Y}_{(s)}; 1, \beta) \\ & \leq T_m(\bar{Y}_{(s)}; 1, [\beta]+1). \end{aligned} \quad (4.3.22)$$

Since $T_m(\bar{Y}_{(s)}; 1, [\beta]) \xrightarrow{a.s.} u_0$ and $T_m(\bar{Y}_{(s)}; 1, [\beta]+1) \xrightarrow{a.s.} u_0$ as $m \rightarrow \infty$ so $T_m(\bar{Y}_{(s)}; 1, \beta) \xrightarrow{a.s.} u_0$. This completes the proof of Lemma 4.3.1.

For a general $\alpha (>1)$, we use Lemma 4.3.1 to prove the following lemma. This lemma plays a vital role in proving Theorem 4.3.1 below.

Lemma 4.3.2. Suppose $\bar{Y}_{(s)} \sim N(X_*b_0, (V + \delta_0)I_m)$. Then $T_m(\bar{Y}_{(s)}; \alpha, \beta)$ as defined in (4.3.11) converges a.s. (as $m \rightarrow \infty$) to u_0 for $\alpha > 1$.

Proof of Lemma 4.3.2. Denote $f_{\alpha, \beta}(u | \bar{Y}_{(s)})$ by $f_{\alpha, \beta}(u)$ for convenience. Then for fixed β and $0 < \alpha < \alpha'$.

$$f_{\alpha', \beta}(u) / f_{\alpha, \beta}(u) \propto (1 - u)^{\alpha' - \alpha} \downarrow \text{ in } u. \quad (4.3.23)$$

Integrating by parts the numerator in (4.3.11), for $\alpha \geq 2$,

$$\begin{aligned} E_{\alpha, \beta}(U | \bar{Y}_{(s)}) &= (m - p + 2\beta) / S_m \\ &\quad - 2(\alpha - 1) S_m^{-1} E_{\alpha, \beta}[U(1-U)^{-1} | \bar{Y}_{(s)}]. \end{aligned} \quad (4.3.24)$$

Using the MLR property as given in (4.3.23), the fact that $u/(1 - u) \uparrow$ in u , and Lemma 2(i) of Lehmann (1986, p. 85) once again, it follows that

$$E_{\alpha, \beta}[U(1-U)^{-1} | \bar{Y}_{(s)}] \leq E_{2, \beta}[U(1-U)^{-1} | \bar{Y}_{(s)}]$$

$$= \left\{ \left[E_{1,\beta}(U|\bar{Y}_{(s)}) \right]^{-1} - 1 \right\}^{-1}$$

$$\xrightarrow{\text{a.s.}} (u_0^{-1} - 1)^{-1} \text{ as } m \rightarrow \infty \quad (4.3.25)$$

from Lemma 4.3.1. Since $S_m/(m-p) \xrightarrow{\text{a.s.}} u_0^{-1}$ as $m \rightarrow \infty$, it follows from (4.3.24), (4.3.25), (4.3.7) and (4.3.11) that

$$T_m(\bar{Y}_{(s)}; \alpha, \beta) = E_{\alpha,\beta}(U|\bar{Y}_{(s)}) \rightarrow u_0 \text{ a.s.} \quad (4.3.26)$$

as $m \rightarrow \infty$ for $\alpha \geq 2$. Thus, for $\alpha \geq 2$, $\lim_{m \rightarrow \infty} T_m(\bar{Y}_{(s)}; \alpha, \beta) = u_0$ a.s. Now, for $1 < \alpha < 2$, using the MLR property once again,

$$E_{2,\beta}(U|\bar{Y}_{(s)}) \leq E_{\alpha,\beta}(U|\bar{Y}_{(s)}) \leq E_{1,\beta}(U|\bar{Y}_{(s)}). \quad (4.3.27)$$

Since both $E_{1,\beta}(U|\bar{Y}_{(s)}) = T_m(\bar{Y}_{(s)}; 1, \beta)$ and $E_{2,\beta}(U|\bar{Y}_{(s)}) = T_m(\bar{Y}_{(s)}; 2, \beta)$ converge a.s. to u_0 as $m \rightarrow \infty$, it follows from (4.3.27) that

$$\begin{aligned} T_m(\bar{Y}_{(s)}; \alpha, \beta) \\ = E_{\alpha,\beta}(U|\bar{Y}_{(s)}) \xrightarrow{\text{a.s.}} u_0 \text{ as } m \rightarrow \infty \text{ for } 1 < \alpha < 2. \end{aligned}$$

Thus $T_m(\bar{Y}_{(s)}; \alpha, \beta) \xrightarrow{\text{a.s.}} u_0$ as $m \rightarrow \infty$ for all $\alpha \geq 1$, and the proof of Lemma 4.3.2 is complete.

Remark 4.3.1. For $0 < \alpha < 1$ we conjecture that the Lemma 4.3.2 is true. This is because

$$\begin{aligned}
T_m(\bar{Y}_{(s)}; \alpha, \beta) \\
&= E_{\alpha, \beta}(U | \bar{Y}_{(s)}) = 1 - E_{\alpha, \beta}(1 - U | \bar{Y}_{(s)}) \\
&= 1 - \left[E_{\alpha+1, \beta}(1 - U)^{-1} | \bar{Y}_{(s)} \right]^{-1},
\end{aligned}$$

$$E_{\alpha+1, \beta} \left[(1 - U)^{-1} | \bar{Y}_{(s)} \right] = 1 + \sum_{i=1}^{\infty} E_{\alpha+1, \beta}(U^i | \bar{Y}_{(s)}).$$

Hence, by Lemma 4.3.2, for any fixed i ,

$$E_{\alpha+1, \beta} \left[U^i | \bar{Y}_{(s)} \right] = \sum_{j=0}^{i-1} E_{\alpha+1, \beta+j}(U | \bar{Y}_{(s)}) \rightarrow u_0^i \text{ a.s. as } m \rightarrow \infty.$$

Hence for each $\ell = 1, 2, \dots$,

$$\lim_{m \rightarrow \infty} \sum_{i=1}^{\ell} E_{\alpha+1, \beta}(U^i | \bar{Y}_{(s)}) = \sum_{i=1}^{\ell} u_0^i \text{ a.s.}$$

Finally making $\ell \rightarrow \infty$ we have

$$\lim_{\ell \rightarrow \infty} \lim_{m \rightarrow \infty} \sum_{i=1}^{\ell} E_{\alpha+1, \beta}(U^i | \bar{Y}_{(s)}) = \frac{u_0}{1 - u_0} \text{ a.s.}$$

Hence, if we can change the order of the above iterated limits, then

$$\lim_{m \rightarrow \infty} \sum_{i=1}^{\infty} E_{\alpha+1, \beta}(U^i | \bar{Y}_{(s)}) = \frac{u_0}{1 - u_0} \text{ a.s.}$$

or equivalently

$$\lim_{m \rightarrow \infty} E_{\alpha, \beta}(U | \bar{Y}_{(s)}) = u_0 \text{ a.s.}$$

However, one has to justify the change in the order of these limits.

We now turn to the main theorem of this section.

Theorem 4.3.1 below proves the A.O. property of $\mathfrak{e}_{BI}$ defined in (4.3.6).

Theorem 4.3.1. Suppose $\bar{Y}_{(s)} \sim N(\bar{X}_* \bar{b}_0, (V + \delta_0) I_m)$. Assume the condition (4.2.2.1). Then, if $\alpha \geq 1$, $r_{Q_m}(\pi_0, \mathfrak{e}_{BI}) - r_{Q_m}(\pi_0, \mathfrak{e}_{SB}^*) \rightarrow 0$ as $m \rightarrow \infty$.

Proof of Theorem 4.3.1. From (4.3.10), we have

$$\begin{aligned}
 & r_{Q_m}(\pi_0, \mathfrak{e}_{BI}) - r_{Q_m}(\pi_0, \mathfrak{e}_{SB}^*) \\
 &= m^{-1} E^* \left[\left(\mathfrak{e}_{BI}(\bar{Y}_{(s)}) - \mathfrak{e}_{SB}^*(\bar{Y}_{(s)}) \right)^T Q_m \left(\mathfrak{e}_{BI}(\bar{Y}_{(s)}) - \mathfrak{e}_{SB}^*(\bar{Y}_{(s)}) \right) \right] \\
 &\leq m^{-1} \text{ch}_L(Q_m) E^* \left\| \mathfrak{e}_{BI}(\bar{Y}_{(s)}) - \mathfrak{e}_{SB}^*(\bar{Y}_{(s)}) \right\|^2, \text{ by Lemma 4.2.2.1} \\
 &= m^{-1} \text{ch}_L(Q_m) E^* \left\| (\bar{Y}_{(s)} - P_{X_*} \bar{Y}_{(s)}) E_{\alpha, \beta}(U | \bar{Y}_{(s)}) - u_0(\bar{Y}_{(s)} - \bar{X}_* \bar{b}_0) \right\|^2, \\
 &\quad \text{by (4.3.6) and (4.3.8)} \\
 &= m^{-1} \text{ch}_L(Q_m) E^* \left\| (\bar{Y}_{(s)} - P_{X_*} \bar{Y}_{(s)}) (E_{\alpha, \beta}(U | \bar{Y}_{(s)}) - u_0) - u_0 (P_{X_*} \bar{Y}_{(s)} - \bar{X}_* \bar{b}_0) \right\|^2 \\
 &= m^{-1} \text{ch}_L(Q_m) \\
 &\quad \times \left[E^* \left\{ \left(E_{\alpha, \beta}(U | \bar{Y}_{(s)}) - u_0 \right)^2 \bar{Y}_{(s)}^T (I_m - P_{X_*}) \bar{Y}_{(s)} \right\} + u_0^2 E^* \left\| P_{X_*} \bar{Y}_{(s)} - \bar{X}_* \bar{b}_0 \right\|^2 \right]
 \end{aligned}$$

(4.3.28)

The last equality follows since $E_{\alpha, \beta}(U|\bar{Y}_{(s)})$ is a function of $\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)}$ and hence $\left(E_{\alpha, \beta}(U|\bar{Y}_{(s)}) - u_0\right)^2 \times \bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)}$ is a function of $\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)}$ which is distributed independently of $P_{X_*}\bar{Y}_{(s)}$ under the distribution $m\pi_0(\bar{Y}_{(s)})$.

Also, since under $m\pi_0(\bar{Y}_{(s)})$, $\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)} \sim (V + \delta_0)\chi_{m-p}^2$, $(m - p)^{-1}\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)} \xrightarrow{a.s.} V + \delta_0$. Hence by Lemma 4.3.2 $\left(E_{\alpha, \beta}(U|\bar{Y}_{(s)}) - u_0\right)^2 m^{-1}\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)} \xrightarrow{a.s.} 0$ as $m \rightarrow \infty$ under $m\pi_0(\bar{Y}_{(s)})$. Also, $\left|E_{\alpha, \beta}(U|\bar{Y}_{(s)}) - u_0\right| \leq 1$ and $m^{-1}\bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)}$ is uniformly integrable in m . Then $m^{-1}E^*\left[\left(E_{\alpha, \beta}(U|\bar{Y}_{(s)}) - u_0\right)^2 \bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)}\right] \rightarrow 0$ as $m \rightarrow \infty$. Also, $E^*[P_{X_*}\bar{Y}_{(s)} - X_*b_0]^2 = (V + \delta_0)p$. Now from (4.3.28) and the condition (4.2.2.1), we have

$$\begin{aligned} 0 &\leq \lim_{m \rightarrow \infty} \left[r_{Q_m}(\pi_0, e_{BI}) - r_{Q_m}(\pi_0, e_{SB}^*) \right] \\ &\leq \overline{\lim}_{m \rightarrow \infty} \left[r_{Q_m}(\pi_0, e_{BI}) - r_{Q_m}(\pi_0, e_{SB}^*) \right] \\ &\leq \Gamma \left[\overline{\lim}_{m \rightarrow \infty} m^{-1}E^* \left\{ \left(E_{\alpha, \beta}(U|\bar{Y}_{(s)}) - u_0 \right)^2 \bar{Y}_{(s)}^T(I_m - P_{X_*})\bar{Y}_{(s)} \right\} \right. \\ &\quad \left. + \overline{\lim}_{m \rightarrow \infty} m^{-1}(V + \delta_0)pu_0^2 \right] = 0. \end{aligned}$$

Therefore $r_{Q_m}(\pi_0, e_{BI}) - r_{Q_m}(\pi_0, e_{SB}^*) \rightarrow 0$ as $m \rightarrow \infty$ and the proof of Theorem 4.3.1 is complete.

Now we will return to the prediction of finite population mean vector γ . From I'-III', it follows that the HB predictor of γ is given by

$$\begin{aligned} \varepsilon_{BF}(\bar{Y}_{(s)}) &= \text{Diag}(1-f_1, \dots, 1-f_m) \bar{Y}_{(s)} \\ &+ \text{Diag}(f_1, \dots, f_m) \varepsilon_{BI}(\bar{Y}_{(s)}). \end{aligned} \quad (4.3.29)$$

And from I' it follows that the subjective Bayes predictor of γ under the prior π_0 is given by

$$\begin{aligned} \varepsilon_{SB}(\bar{Y}_{(s)}) &= \text{Diag}(1-f_1, \dots, 1-f_m) \bar{Y}_{(s)} \\ &+ \text{Diag}(f_1, \dots, f_m) \varepsilon_{SB}^*(\bar{Y}_{(s)}). \end{aligned} \quad (4.3.30)$$

We will now show the asymptotic optimality of the HB predictor $\varepsilon_{BF}(\bar{Y}_{(s)})$. From (4.2.1.1),

$$\begin{aligned} r_{Q_m}(\pi_0, \varepsilon_{BF}) - r_{Q_m}(\pi_0, \varepsilon_{SB}) &= m^{-1} E^* \left[\left\{ \text{Diag}(f_1, \dots, f_m) \left(\varepsilon_{BI}(\bar{Y}_{(s)}) - \varepsilon_{SB}^*(\bar{Y}_{(s)}) \right) \right\}^T Q_m \right. \\ &\quad \times \left. \left\{ \text{Diag}(f_1, \dots, f_m) \left(\varepsilon_{BI}(\bar{Y}_{(s)}) - \varepsilon_{SB}^*(\bar{Y}_{(s)}) \right) \right\} \right] \\ &\leq m^{-1} \text{ch}_L(Q_m) \text{ch}_L(\text{Diag}(f_1^2, \dots, f_m^2)) \\ &\quad \times E^* \left[\left\| \varepsilon_{BI}(\bar{Y}_{(s)}) - \varepsilon_{SB}^*(\bar{Y}_{(s)}) \right\|^2 \right] \quad (\text{by Lemma 4.2.1.1}) \\ &\leq m^{-1} \text{ch}_L(Q_m) E^* \left[\left\| \varepsilon_{BI}(\bar{Y}_{(s)}) - \varepsilon_{SB}^*(\bar{Y}_{(s)}) \right\|^2 \right] \end{aligned}$$

since $f_i^2 \leq 1$ for all $i = 1, \dots, m$. From this we have as in Theorem 4.3.1 that $r_{Q_m}(\pi_0, \epsilon_{BF}) - r_{Q_m}(\pi_0, \epsilon_{SB}) \rightarrow 0$ as $m \rightarrow \infty$.

4.4 Asymptotic Optimality with Unknown Variance Components

In this section, we dispense with the assumption of a known variance component at the first stage, and consider a hierarchical model similar to the one considered in Section 4.3.

- I. Conditional on $\Theta = \theta$, $B = b$, $R = r$ and $\Lambda = \lambda$, $Y_i \sim N(\theta_i 1_{N_i}, r^{-1} I_{N_i})$, $i = 1, \dots, m$ independently;
- II. conditional on $B = b$, $R = r$ and $\Lambda = \lambda$, $\Theta \sim N(X_* b, (\lambda r)^{-1} I_m)$, where $X_*(m \times p)$ is assumed to have rank $p < m$;
- III. B , R and $\epsilon = \Lambda R$ are mutually independently distributed with $B \sim \text{uniform}(R^p)$, ϵ has pdf $f(\epsilon) \propto \epsilon^{-\alpha-1} \left(0 < \alpha < \frac{1}{2}(m-p)\right)$ and $R \sim \text{gamma}\left(\frac{1}{2}a_0, \frac{1}{2}g_0\right)$, that is R has pdf $h(r) \propto \exp\left(-\frac{1}{2}a_0 r\right) r^{\frac{1}{2}g_0-1}$, where $a_0 \geq 0$, and g_0 (real) satisfies $(n-1)m + g_0 > 0$ where n is the sample size from each stratum. Thus, ϵ has improper gamma pdf, while R has a proper or an improper gamma pdf with its parameters satisfying certain conditions.

Here also we are interested in proving the A.O. property of

the HB predictor of γ . We will attain this by proving the asymptotic optimality of the HB predictor of θ . Assume that we have a sample $Y_{i1}, \dots, Y_{in}$ of size n from the i^{th} stratum. Let $\bar{Y}_{(s)} = (\bar{Y}_{1(s)}, \dots, \bar{Y}_{m(s)})^T$ where $\bar{Y}_{i(s)} = n^{-1} \sum_{j=1}^n Y_{ij}$, $i = 1, \dots, m$ and $S = \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i(s)})^2$. From I, it easily follows that $(\bar{Y}_{(s)}^T, S)^T$ is minimal sufficient for $(\theta^T, r)^T$. So our posterior distribution will depend only on $(\bar{Y}_{(s)}^T, S)^T$.

Using I - III with routine calculations, one obtains the following results:

(i) conditional on $\bar{Y}_{(s)} = \bar{y}_{(s)}$, $S = s$, $R = r$ and $\Lambda = \lambda$,

$$\bar{\theta} \sim N \left[(n+\lambda)^{-1} (n\bar{y}_{(s)} + \lambda P_{X_*} \bar{y}_{(s)}), (nr)^{-1} (n+\lambda)^{-1} (nI_m + \lambda P_{X_*}) \right]; \quad (4.4.1)$$

(ii) conditional on $\bar{Y}_{(s)} = \bar{y}_{(s)}$, $S = s$ and $\Lambda = \lambda$,

$$R \sim \text{Gamma} \left(\frac{1}{2} (a_0 + s + \frac{n\lambda}{n+\lambda} \bar{y}_{(s)}^T (I_m - P_{X_*}) \bar{y}_{(s)}), \frac{1}{2} (nm - p - 2\alpha + g_0) \right);$$

(iii) conditional on $\bar{Y}_{(s)} = \bar{y}_{(s)}$ and $S = s$, Λ has pdf

$$\begin{aligned} f(\lambda | \bar{y}_{(s)}, s) &\propto (\lambda / (n+\lambda))^{\frac{1}{2}(m-p)} \lambda^{-\alpha-1} \\ &\times \left[a_0 + s + \frac{n\lambda}{n+\lambda} \bar{y}_{(s)}^T (I_m - P_{X_*}) \bar{y}_{(s)} \right]^{-\frac{1}{2}(nm-p-2\alpha+g_0)}. \end{aligned} \quad (4.4.2)$$

Let $U = \Lambda / (n + \Lambda)$. Then, from (4.4.2) it follows that the conditional pdf of U given $\bar{Y}_{(s)} = \bar{y}_{(s)}$ and $S = s$ is

$$f_{\alpha}(u | \bar{y}_{(s)}, s) \propto u^{\frac{1}{2}(m-p-2\alpha)-1} (1-u)^{\alpha-1} (1+uF)^{-\frac{1}{2}(nm-p-2\alpha+g_0)}, \quad (4.4.3)$$

$0 < u < 1$, where $F = n\bar{y}_{(s)}^T (I_m - P_{X_*}) \bar{y}_{(s)} / (a_0 + s)$. Note that if $a_0 = 0$, F is a multiple of a usual F statistic. Also, from (4.4.1) and (4.4.3), one gets the HB estimator of θ under the loss (4.2.2) given by

$$\begin{aligned} \tilde{\theta}_{HB}(\bar{Y}_{(s)}, S) &= E[\bar{\theta} | \bar{Y}_{(s)}, S] \\ &= \left[1 - E_{\alpha}(U | \bar{Y}_{(s)}, S) \right] \bar{Y}_{(s)} + E_{\alpha}(U | \bar{Y}_{(s)}, S) P_{X_*} \bar{Y}_{(s)} \end{aligned} \quad (4.4.4)$$

where

$$\begin{aligned} E_{\alpha}(U | \bar{Y}_{(s)}, S) &= \int_0^1 u^{\frac{1}{2}(m-p-2\alpha)} (1-u)^{\alpha-1} (1+uF)^{-\frac{1}{2}(nm-p-2\alpha+g_0)} du \\ &\div \int_0^1 u^{\frac{1}{2}(m-p-2\alpha)-1} (1-u)^{\alpha-1} (1+uF)^{-\frac{1}{2}(nm-p-2\alpha+g_0)} du. \end{aligned} \quad (4.4.5)$$

We now examine the asymptotic (as $m \rightarrow \infty$) behavior of $\tilde{\epsilon}_{HB}$ as an estimator of θ under the subjective prior π_0 which specifies the value r_0 for r and the distribution $\theta \sim N(\underline{x}^* \underline{b}_0, (\lambda_0 r_0)^{-1} \underline{I}_m)$, and under the loss $L(\theta, \underline{a})$ given in (4.2.2). Then, marginally $\bar{Y}_{(S)}$ and S are mutually independent with $\bar{Y}_{(S)} \sim N(\underline{x}^* \underline{b}_0, r_0^{-1}(n^{-1} + \lambda_0^{-1}) \underline{I}_m)$ and $S \sim r_0^{-1} \chi_{(n-1)m}^2$. We prove a series of lemmas culminating eventually in Theorem 4.4.1 which establishes the A.O. property of $\tilde{\epsilon}_{HB}$ as an estimator of θ under the general quadratic loss given in (4.2.2).

Denote the expectation $E_\alpha(U | \bar{Y}_{(S)}, S)$ given in (4.4.5) by $W_m(\bar{Y}_{(S)}, S; \alpha)$. We subscripted it by m so tht we can consider the sequence $\{W_m(\bar{Y}_{(S)}, S; \alpha): m \geq 1\}$.

First the following lemma is proved. Then this lemma will be used to prove a general lemma for $\alpha \geq 1$.

Lemma 4.4.1. Let $u_0 = \lambda_0 / (n + \lambda_0)$. Consider the case $\alpha = 1$. Then, as $m \rightarrow \infty$, $W_m(\bar{Y}_{(S)}, S; 1) \xrightarrow{a.s.} u_0$, where $\bar{Y}_{(S)}$ and S are mutually independent with $\bar{Y}_{(S)} \sim N(\underline{x}^* \underline{b}_0, r_0^{-1}(n^{-1} + \lambda_0^{-1}) \underline{I}_m)$ and $S \sim r_0^{-1} \chi_{(n-1)m}^2$.

Proof of Lemma 4.4.1. Assume m is so large that $t_1 = \frac{1}{2}(m - p - 2) \geq 1$ and $t_2 = \frac{1}{2}((n - 1)m + g_0) > 1$. Since $\alpha = 1$, it follows from (4.4.5) that

$$\begin{aligned}
& W_m(\bar{Y}_{(s)}, S; 1) \\
&= \int_0^1 u^{t_1} (1 + uF)^{-(t_1+t_2)} du \Big/ \int_0^1 u^{t_1-1} (1 + uF)^{-(t_1+t_2)} du \\
&= F^{-1} \int_0^1 \left(uF/(1 + uF) \right)^{t_1} \left(1/(1 + uF) \right)^{t_2-2} \left(F/(1 + uF)^2 \right) du \\
&\quad \div \int_0^1 \left(uF/(1 + uF) \right)^{t_1-1} \left(1/(1 + uF) \right)^{t_2-1} \left(F/(1 + uF)^2 \right) du \\
&= F^{-1} \int_0^{F/(1+F)} v^{t_1} (1-v)^{t_2-2} dv \Big/ \int_0^{F/(1+F)} v^{t_1-1} (1-v)^{t_2-1} dv. \\
&\hspace{25em} (4.4.6)
\end{aligned}$$

Integration by parts gives

$$\begin{aligned}
& \int_0^{F/(1+F)} v^{t_1} (1-v)^{t_2-2} dv \\
&= -\left(F/(1 + F) \right)^{t_1} \left(1/(1 + F) \right)^{t_2-1} (t_2 - 1)^{-1} \\
&\quad + \left(t_1/(t_2 - 1) \right) \int_0^{F/(1+F)} v^{t_1-1} (1-v)^{t_2-1} dv. \quad (4.4.7)
\end{aligned}$$

Combining (4.4.6) and (4.4.7), one gets

$$\begin{aligned}
& W_m(\bar{Y}_{(s)}, S; 1) \\
&= t_1 / \left((t_2 - 1)F \right) - (t_2 - 1)^{-1} F^{t_1-1} (1 + F)^{-(t_1+t_2-1)} \\
&\quad \times \left[\int_0^{F/(1+F)} v^{t_1-1} (1-v)^{t_2-1} dv \right]^{-1}. \quad (4.4.8)
\end{aligned}$$

Note that since $S \sim r_0^{-1} \chi_{(n-1)m}^2$, $S/((n-1)m) \xrightarrow{a.s.} r_0^{-1}$ as $m \rightarrow \infty$. Again, using the idempotency of $I_m - P_{X_*}$, it follows that $n \bar{Y}_{(s)}^T (I_m - P_{X_*}) \bar{Y}_{(s)} = n (\bar{Y}_{(s)} - X_* b_0)^T (I_m - P_{X_*}) (\bar{Y}_{(s)} - X_* b_0) \sim r_0^{-1} u_0^{-1} \chi_{m-p}^2$. Hence, $n \bar{Y}_{(s)}^T (I_m - P_{X_*}) \bar{Y}_{(s)} / m \xrightarrow{a.s.} r_0^{-1} u_0^{-1}$ as $m \rightarrow \infty$. Thus, $t_1 / ((t_2 - 1)F) = \{(m - p - 2) / ((n - 1)m + g_0 - 2)\} F^{-1} \xrightarrow{a.s.} (n - 1)^{-1} \times (n - 1) r_0^{-1} / (r_0^{-1} u_0^{-1}) = u_0$ as $m \rightarrow \infty$. Hence, for proving Lemma 4.4.1, it suffices to show that the second term in the rhs of (4.4.8) converges to zero a.s. as $m \rightarrow \infty$ under the distribution of $\bar{Y}_{(s)}$ and S given in the lemma.

With this end, assume first that t_1 is a positive integer, and use successive integration by parts to obtain

$$\begin{aligned} & \int_0^{F/(1+F)} v^{t_1-1} (1-v)^{t_2-1} dv \\ &= -(1+F)^{-(t_1+t_2-1)} \sum_{j=0}^{t_1-1} \left\{ (t_1-1)_j F^{t_1-j-1} \right. \\ & \quad \left. \div (t_2 \dots (t_2+j)) \right\} + (t_1-1)! \Gamma(t_2) / \Gamma(t_1+t_2), \end{aligned} \quad (4.4.9)$$

where $(x)_0 = 1$, $(x)_j = x(x-1) \dots (x-j+1)$ for $j = 1, 2, \dots, x$ and $\Gamma(\cdot)$ is the usual gamma function. Hence, from (4.4.9),

$$\begin{aligned}
& (t_2 - 1)(1 + F)^{t_1+t_2-1} F^{-(t_1-1)} \int_0^{F/(1+F)} v^{t_1-1} (1-v)^{t_2-1} dv \\
&= -(t_2 - 1) \sum_{j=0}^{t_1-1} \frac{(t_1-1)_j F^{-j}}{t_2(t_2+1) \dots (t_2+j)} + (1 + F)^{t_1+t_2-1} \\
&\quad \times F^{-(t_1-1)} (t_2 - 1)(t_1 - 1)! \Gamma(t_2) / \Gamma(t_1 + t_2) \\
&\geq -(t_2 - 1) \sum_{j=0}^{t_1-1} \frac{(t_1-1)_j}{t_2^{j+1}} F^{-j} + (1 + F)^{t_1+t_2-1} \\
&\quad \times F^{-(t_1-1)} (t_2 - 1)(t_1 - 1)! \Gamma(t_2) / \Gamma(t_1 + t_2) \\
&\geq -(t_2 - 1) t_2^{-1} \sum_{j=0}^{\infty} \left\{ (t_1 - 1) / t_2 \right\}^j F^{-j} + (1 + F)^{t_1+t_2-1} \\
&\quad \times F^{-(t_1-1)} (t_2 - 1)(t_1 - 1)! \Gamma(t_2) / \Gamma(t_1 + t_2).
\end{aligned} \tag{4.4.10}$$

Note that the first term in the rhs of (4.4.10) converges a.s. to $-\sum_{j=0}^{\infty} (n-1)^{-j} ((n-1)u_0)^j = -\sum_{j=0}^{\infty} u_0^j = -(1-u_0)^{-1}$ as $m \rightarrow \infty$. Also, using Stirling's bounds for gamma functions,

second term in the rhs of (4.4.10)

$$\begin{aligned}
&\geq (1 + F)^{t_1+t_2-1} F^{-(t_1-1)} (2\pi)^{\frac{1}{2}} (t_2 - 1) \\
&\quad \times \exp(-(t_1 - 1)) (t_1 - 1)^{t_1 - \frac{1}{2}} \exp(-(t_2 - 1)) (t_2 - 1)^{t_2 - \frac{1}{2}}
\end{aligned}$$

$$\begin{aligned}
& \div \left[\exp(-(t_1+t_2-1))(t_1+t_2-1)^{t_1+t_2-\frac{1}{2}} \exp\left(1/(12(t_1+t_2-1))\right) \right] \\
& = (1+F)^{t_1+t_2-1} F^{-(t_1-1)} (2\pi)^{\frac{1}{2}} (t_2-1) e(t_1-1)^{t_1-\frac{1}{2}} (t_2-1)^{t_2-\frac{1}{2}} \\
& \quad \times (t_1+t_2-1)^{-(t_1+t_2-\frac{1}{2})} \exp\left(-\frac{1}{12(t_1+t_2-1)}\right). \tag{4.4.11}
\end{aligned}$$

Recall that $t_1 = \frac{1}{2}(m-p-2)$, $t_2 = \frac{1}{2}((n-1)m + g_0)$, and $F \xrightarrow{\text{a.s.}} (n-1)^{-1} u_0^{-1}$ as $m \rightarrow \infty$. Write $h(m)$ for the logarithm of the rhs of (4.4.11). Then,

$$\begin{aligned}
& \lim_{m \rightarrow \infty} m^{-1} h(m) \\
& \stackrel{\text{a.s.}}{\geq} \frac{1}{2} n \log\left(1 + \frac{u_0^{-1}}{n-1}\right) - \frac{1}{2} \log(u_0^{-1}/(n-1)) \\
& \quad + \frac{1}{2}(n-1) \log(n-1) - \frac{1}{2} n \log n \\
& = \frac{1}{2} \left[n \log(n-1 + u_0^{-1}) - \log u_0^{-1} - n \log n \right] \\
& = \frac{1}{2} \left[n \log\left(1 + \frac{1 + u_0^{-1}}{n}\right) - \log u_0^{-1} \right] \\
& = \frac{1}{2} \log \left[\frac{\left(1 + \frac{1}{\lambda_0}\right)^n}{\left(1 + \frac{n}{\lambda_0}\right)} \right] > 0, \tag{4.4.12}
\end{aligned}$$

using $(1+x)^n > 1+nx$ for $0 < x$. Hence $h(m) \rightarrow \infty$ a.s. as $m \rightarrow \infty$, and it follows from (4.4.11) and (4.4.12)

that the second term in the rhs of (4.4.8) converges to zero a.s. as $m \rightarrow \infty$.

If t_1 is not an integer, write

$$h_{t_1}(u) = u^{t_1-1} (1 + uF)^{-(t_1+t_2)} \\ \div \int_0^1 u^{t_1-1} (1 + uF)^{-(t_1+t_2)} du.$$

Then, for $t_1 < t'_1$, $h_{t'_1}(u)/h_{t_1}(u) \propto (u/(1 + uF))^{t'_1-t_1} \uparrow$ in u .

Using this MLR property, and writing $W_m(\bar{Y}_{(s)}, S; 1)$ as

$E_{t_1}^*(U|\bar{Y}_{(s)}, S)$, it follows that $E_{t_1}^*(U|\bar{Y}_{(s)}, S)$ is $\uparrow$ in t_1

where $E_{t_1}^*$ denotes the expectation w.r.t. $h_{t_1}(u)$. Hence, if

$[t_1]$ denotes the integer part of t_1 , since $[t_1] \geq 1$,

$$E_{[t_1]}^*(U|\bar{Y}_{(s)}, S) \leq E_{t_1}^*(U|\bar{Y}_{(s)}, S) \leq E_{[t_1]+1}^*(U|\bar{Y}_{(s)}, S).$$

Since both $E_{[t_1]}^*(U|\bar{Y}_{(s)}, S)$ and $E_{[t_1]+1}^*(U|\bar{Y}_{(s)}, S)$ converge

a.s. to u_0 as $m \rightarrow \infty$, it follows that $E_{t_1}^*(U|\bar{Y}_{(s)}, S) \xrightarrow{a.s.} u_0$

as $m \rightarrow \infty$. The proof of Lemma 4.4.1 is complete.

Lemma 4.4.2. Consider the set up of Lemma 4.4.1 with

$\alpha \geq 1$. Then $W_m(\bar{Y}_{(s)}, S; \alpha) \xrightarrow{a.s.} u_0$ as $m \rightarrow \infty$.

Proof of Lemma 4.4.2. Assume m is large enough so that t_3

$= \frac{1}{2}(m - p - 2\alpha) > 0$. Now,

$$\begin{aligned}
& w_m(\bar{Y}_{(s)}, S; \alpha) \\
&= \int_0^1 u^{t_3} (1-u)^{\alpha-1} (1+uF)^{-(t_3+t_2)} du \\
&\quad + \int_0^1 u^{t_3-1} (1-u)^{\alpha-1} (1+uF)^{-(t_3+t_2)} du. \quad (4.4.13)
\end{aligned}$$

Consider first the case when $\alpha \geq 2$. Integration by parts gives

$$\begin{aligned}
& \int_0^1 u^{t_3} (1-u)^{\alpha-1} (1+uF)^{-(t_3+t_2)} du \\
&= \left\{ t_3 / (t_3 + t_2 - 1) \right\} F^{-1} \int_0^1 u^{t_3-1} (1-u)^{\alpha-1} \\
&\quad \times (1+uF)^{-(t_3+t_2)+1} du - \left\{ (\alpha-1) / (t_3 + t_2 - 1) \right\} F^{-1} \\
&\quad \times \int_0^1 u^{t_3} (1-u)^{\alpha-2} (1+uF)^{-(t_3+t_2)+1} du \\
&= \left\{ t_3 / (t_3 + t_2 - 1) \right\} F^{-1} \int_0^1 u^{t_3-1} (1-u)^{\alpha-1} \\
&\quad \times (1+uF)^{-(t_3+t_2)} du + \left\{ t_3 / (t_3 + t_2 - 1) \right\} \\
&\quad \times \int_0^1 u^{t_3} (1-u)^{\alpha-1} (1+uF)^{-(t_3+t_2)} du
\end{aligned}$$

$$\begin{aligned}
& - \{(\alpha - 1)/(t_3 + t_2 - 1)\} F^{-1} \\
& \times \int_0^1 u^{t_3} (1 - u)^{\alpha-2} (1 + uF)^{-(t_3+t_2)+1} du. \quad (4.4.14)
\end{aligned}$$

Combining (4.4.13) and (4.4.14) one gets

$$\begin{aligned}
& W_m(\bar{Y}_{(S)}, S; \alpha) \\
& = (t_3/(t_2 - 1)) F^{-1} - \{(\alpha - 1)/(t_2 - 1)\} F^{-1} \\
& \times E_\alpha[U(1 + UF)/(1 - U) | \bar{Y}_{(S)}, S] \quad (4.4.15)
\end{aligned}$$

where recall that E_α is the expectation w.r.t. f_α given in (4.4.3). Hence, from (4.4.15),

$$\begin{aligned}
& W_m(\bar{Y}_{(S)}, S; \alpha) \\
& \leq (t_3/(t_2 - 1)) F^{-1} \xrightarrow{a.s.} u_0 \text{ as } m \rightarrow \infty. \quad (4.4.16)
\end{aligned}$$

Also, writing $f_\alpha(u | \bar{Y}_{(S)}, S)$ as $f_\alpha(u)$ for $\alpha < \alpha'$,

$f_{\alpha'}(u)/f_\alpha(u) \propto [(u^{-1} + F)(1 - u)]^{\alpha' - \alpha} \downarrow$ in u . Hence, since $u/(1 - u)$ is $\uparrow$ in u , for $\alpha \geq 2$,

$$\begin{aligned}
& E_\alpha[U/(1 - U) | \bar{Y}_{(S)}, S] \\
& \leq E_2[U/(1 - U) | \bar{Y}_{(S)}, S] \\
& = \int_0^1 u^{t_3} (1 + uF)^{-(t_3+t_2)} du
\end{aligned}$$

$$\begin{aligned}
& \div \int_0^1 u t_3^{-1} (1 - u) (1 + uF)^{-(t_3+t_2)} du \\
& \leq \int_0^1 u t_3^{-1} (1 + uF)^{-(t_3+t_2)} du \\
& \div \int_0^1 u t_3^{-1} (1 - u) (1 + uF)^{-(t_3+t_2)} du \\
& = \left[E_1((1 - U)|\bar{Y}_{(S)}, S) \right]^{-1} \xrightarrow{a.s.} (1 - u_0)^{-1} \text{ as } m \rightarrow \infty.
\end{aligned} \tag{4.4.17}$$

From (4.4.15) and (4.4.17),

$$\begin{aligned}
& W_m(U|\bar{Y}_{(S)}, S; \alpha) \\
& \geq (t_3/(t_2 - 1))F^{-1} - \{(\alpha - 1)/(t_2 - 1)\}F^{-1} \\
& \quad \times (1 + F) \left[E_1\{(1 - U)|\bar{Y}_{(S)}, S\} \right]^{-1} \\
& \xrightarrow{a.s.} u_0 - 0 = u_0 \text{ as } m \rightarrow \infty.
\end{aligned} \tag{4.4.18}$$

It follows from (4.4.16) and (4.4.18) that for $\alpha \geq 2$

$W_m(\bar{Y}_{(S)}, S; \alpha) \xrightarrow{a.s.} u_0$ as $m \rightarrow \infty$. For $1 < \alpha < 2$, use the inequality $E_2(U|\bar{Y}_{(S)}, S) \leq E_\alpha(U|\bar{Y}_{(S)}, S) \leq E_1(U|\bar{Y}_{(S)}, S)$ and use the fact that both $E_1(U|\bar{Y}_{(S)}, S) = W_m(\bar{Y}_{(S)}, S; 1)$ and $E_2(U|\bar{Y}_{(S)}, S) = W_m(\bar{Y}_{(S)}, S; 2)$ converge a.s. to u_0 as $m \rightarrow \infty$ to conclude that $W_m(\bar{Y}_{(S)}, S; \alpha) \xrightarrow{a.s.} u_0$ as $m \rightarrow \infty$. The proof of Lemma 4.4.2 is complete.

Under the subjective prior π_0 which specifies $r = r_0$, $\lambda = \lambda_0$ and $\Theta \sim N(X_*b_0, (r_0\lambda_0)^{-1}I_m)$ the subjective Bayes estimator of θ is given by

$$\tilde{e}_{SB}(\bar{Y}_{(s)}, S) = (1 - u_0)\bar{Y}_{(s)} + u_0X_*b_0. \quad (4.4.19)$$

The following theorem proves the A.O. property of $\tilde{e}_{HB}$.

Theorem 4.4.1. Assume the condition (4.2.2.1) holds. Then for the prior π_0 given at the beginning of the section and for $\alpha \geq 1$, $r_{Q_m}(\pi_0, \tilde{e}_{HB}) - r_{Q_m}(\pi_0, \tilde{e}_{SB}) \rightarrow 0$ as $m \rightarrow \infty$.

Proof of Theorem 4.4.1. Let E^* denote the expectation taken w.r.t. the joint distribution of $\bar{Y}_{(s)}$ and S specified by the prior π_0 . Then as in (4.3.10), we have

$$\begin{aligned} 0 &\leq r_{Q_m}(\pi_0, \tilde{e}_{HB}) - r_{Q_m}(\pi_0, \tilde{e}_{SB}) \\ &= m^{-1}E^*\left[\left(\tilde{e}_{HB}(\bar{Y}_{(s)}, S) - \tilde{e}_{SB}(\bar{Y}_{(s)}, S)\right)^T Q_m\left(\tilde{e}_{HB}(\bar{Y}_{(s)}, S) - \tilde{e}_{SB}(\bar{Y}_{(s)}, S)\right)\right] \\ &\leq m^{-1}ch_L(Q_m)E^*\left\|\tilde{e}_{HB}(\bar{Y}_{(s)}, S) - \tilde{e}_{SB}(\bar{Y}_{(s)}, S)\right\|^2, \text{ by Lemma 4.2.2.1} \\ &= m^{-1}ch_L(Q_m)E^*\left\|\left(\bar{Y}_{(s)} - P_{X_*}\bar{Y}_{(s)}\right)E_\alpha(U|\bar{Y}_{(s)}, S) - \left(\bar{Y}_{(s)} - X_*b_0\right)u_0\right\|^2 \\ &= m^{-1}ch_L(Q_m) \\ &\quad \times E^*\left\|\left(\bar{Y}_{(s)} - P_{X_*}\bar{Y}_{(s)}\right)\left(E_\alpha(U|\bar{Y}_{(s)}, S) - u_0\right) - u_0\left(P_{X_*}\bar{Y}_{(s)} - X_*b_0\right)\right\|^2. \end{aligned} \quad (4.4.20)$$

Note that $\text{Cov}_{\pi_0}\left((\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)}, P_{X_*}\bar{Y}_{(s)}\right) = r_0^{-1}(n^{-1} + \lambda_0^{-1}) \times (\mathbb{I}_m - P_{X_*})P_{X_*} = r_0^{-1}(n^{-1} + \lambda_0^{-1})(P_{X_*} - P_{X_*}^2) = 0$ since P_{X_*} is idempotent. So under π_0 , $P_{X_*}\bar{Y}_{(s)}$ and $(\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)}$ are statistically independent. On the other hand, from (4.4.3) it follows that $E_\alpha(U|\bar{Y}_{(s)}, S)$ is a function of $(\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)}$ (since $\bar{Y}_{(s)}^T(\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)} = ((\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)})^T(\mathbb{I}_m - P_{X_*})\bar{Y}_{(s)}$) and S . Also $\bar{Y}_{(s)}$ and S are statistically independent. So the two terms inside $\|\cdot\|^2$ in (4.4.20) are independently distributed. From this discussion and noting that $E^*(P_{X_*}\bar{Y}_{(s)}) = X_*b_0$, we have

the rhs of (4.4.20)

$$\begin{aligned}
 &= m^{-1} \text{ch}_L(Q_m) \left[E^* \left\| (\bar{Y}_{(s)} - P_{X_*}\bar{Y}_{(s)}) (E_\alpha(U|\bar{Y}_{(s)}, S) - u_0) \right\|^2 \right. \\
 &\quad \left. + E^* \left\| u_0 (P_{X_*}\bar{Y}_{(s)} - X_*b_0) \right\|^2 \right] \\
 &= m^{-1} \text{ch}_L(Q_m) \left[E^* \left((E_\alpha(U|\bar{Y}_{(s)}, S) - u_0)^2 \bar{Y}_{(s)}^T (\mathbb{I}_m - P_{X_*}) \bar{Y}_{(s)} \right) \right. \\
 &\quad \left. + u_0^2 r_0^{-1} (n^{-1} + \lambda_0^{-1}) \text{tr}(P_{X_*}^2) \right] \\
 &= m^{-1} \text{ch}_L(Q_m) \left[E^* \left((E_\alpha(U|\bar{Y}_{(s)}, S) - u_0)^2 \bar{Y}_{(s)}^T (\mathbb{I}_m - P_{X_*}) \bar{Y}_{(s)} \right) \right. \\
 &\quad \left. + u_0^2 r_0^{-1} n^{-1} p \right] \tag{4.4.21}
 \end{aligned}$$

since $\text{tr}(\mathbf{P}_{\underline{X}_*}^2) = \text{tr}(\mathbf{P}_{\underline{X}_*}) = \text{rank}(\underline{X}_*) = p$. As in the proof of Theorem 4.3.1, we can prove $m^{-1}E^*\left((E_\alpha(U|\bar{Y}_{(s)}), S) - u_0\right)^2 \bar{Y}_{(s)}^T \times (\mathbf{I}_m - \mathbf{P}_{\underline{X}_*}) \bar{Y}_{(s)} \rightarrow 0$ as $m \rightarrow \infty$. Hence from (4.4.20), (4.4.21) and (4.2.2.1), we have

$$\begin{aligned} & \overline{\lim}_{m \rightarrow \infty} \left[r_{\underline{Q}_m}(\pi_0, \tilde{e}_{\text{HB}}) - r_{\underline{Q}_m}(\pi_0, \tilde{e}_{\text{SB}}) \right] \\ & \leq \Gamma \lim_{m \rightarrow \infty} E^* \left((E_\alpha(U|\bar{Y}_{(s)}), S) - u_0 \right)^2 \bar{Y}_{(s)}^T (\mathbf{I}_m - \mathbf{P}_{\underline{X}_*}) \bar{Y}_{(s)} / m \\ & = 0. \end{aligned}$$

Hence, $r_{\underline{Q}_m}(\pi_0, \tilde{e}_{\text{HB}}) - r_{\underline{Q}_m}(\pi_0, \tilde{e}_{\text{SB}}) \rightarrow 0$ as $m \rightarrow \infty$.

We may now apply this theorem for HB prediction of γ . From I - III, it follows that the HB predictor of γ is given by

$$\begin{aligned} \tilde{e}_{\text{BF}}(\bar{Y}_{(s)}, S) &= \text{Diag}(1-f_1, \dots, 1-f_m) \bar{Y}_{(s)} \\ &+ \text{Diag}(f_1, \dots, f_m) \tilde{e}_{\text{HB}}(\bar{Y}_{(s)}, S) \end{aligned} \quad (4.4.22)$$

where $f_i = 1 - n/N_i$, $i = 1, \dots, m$. From I it follows that the subjective Bayes predictor of γ under the prior π_0 is given by

$$\begin{aligned} \tilde{e}_{\text{SB}}(\bar{Y}_{(s)}, S) &= \text{Diag}(1-f_1, \dots, 1-f_m) \bar{Y}_{(s)} \\ &+ \text{Diag}(f_1, \dots, f_m) \tilde{e}_{\text{SB}}(\bar{Y}_{(s)}, S). \end{aligned} \quad (4.4.23)$$

Note that

$$\begin{aligned} \tilde{e}_{BF}(\bar{Y}_{(s)}, S) - e_{SB}(\bar{Y}_{(s)}, S) \\ = \text{Diag}(f_1, \dots, f_m) \left(\tilde{e}_{HB}(\bar{Y}_{(s)}, S) - \tilde{e}_{SB}(\bar{Y}_{(s)}, S) \right) \end{aligned}$$

gives

$$\begin{aligned} \left| \tilde{e}_{BF}(\bar{Y}_{(s)}, S) - e_{SB}(\bar{Y}_{(s)}, S) \right|^2 \\ \leq \left| \tilde{e}_{HB}(\bar{Y}_{(s)}, S) - \tilde{e}_{SB}(\bar{Y}_{(s)}, S) \right|^2. \end{aligned}$$

Using the above inequality, we can easily show as in

Section 4.3 that $r_{Q_m}(\pi_0, \tilde{e}_{BF}) - r_{Q_m}(\pi_0, e_{SB}) \rightarrow 0$ as $m \rightarrow \infty$.

CHAPTER FIVE SIMULTANEOUS BAYESIAN ESTIMATION OF SMALL AREA VARIANCES

5.1 Introduction

In this chapter, we are interested in simultaneous HB estimation of variances from several small areas, where each small area has a finite number of units. This chapter is developed following in part the outlines of the preceding three chapters. We will develop the HB estimator of the finite population variance vector assuming an underlying normal linear model for the superpopulation under a quadratic loss.

We use the notations introduced earlier. Let $\underline{\rho} = (\rho_1, \dots, \rho_m)^T$ denote the finite population variance vector where $\rho_i = (N_i - 1)^{-1} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2$, $i = 1, \dots, m$. We want to find the HB estimator of $\underline{\rho}$ and study its asymptotic optimal property under the loss

$$L(\underline{a}, \underline{\rho}) = m^{-1}(\underline{a} - \underline{\rho})^T \underline{Q}_m (\underline{a} - \underline{\rho}) \quad (5.1.1)$$

where $\underline{Q}_m$ ($m \times m$) is a known n.n.d. and nonnull matrix. This loss has been used in the previous chapter.

Bayesian estimation of variances from stratified samples were considered earlier by Ghosh and Lahiri (1987b) and Lahiri and Tiwari (in press) without incorporating any auxiliary information. Hartley and Rao (1968) introduced auxiliary information in the estimation of variances for stratified samples only in a very special set up. The present chapter treats the variance estimation problem in the general framework of Chapter Two.

In Section 5.2, we have developed under the set up of Section 3.2 (i.e., with known ratios of variance components) the HB estimator of a quadratic form. We have used this result to derive explicitly the HB estimator of ρ . This estimator is considered in greater details in Section 5.3, in the special case of nested error regression model (2.2.3) with $x_{ij} = x_i$, $j = 1, \dots, N_i$; $i = 1, \dots, m$. As in Chapter Four, we study the asymptotic optimality of the proposed estimator. Since this problem is much more algebraically involved than the one we considered for the means in Chapter Four, we restrict ourselves to this special case. While we prove that this HB estimator is asymptotically optimal under the loss (5.1.1) and the subjective prior π_0 of Section 4.4, we establish that the usual sample variance vector turns out to be nonoptimal. To prove these results, we do not need the n_i to be equal.

So far we have assumed that the ratio of variance components is known. In Section 5.4, for this special case again, we derive the HB predictor following the hierarchical Bayes set up of Section 4.4 when both variance components are unknown and are assigned gamma priors (proper or improper). Following the arguments of Section 4.4, we have been able to prove the asymptotic optimality of our HB predictor under the additional assumption that all the n_i are equal. Ghosh and Lahiri (1987b) and Lahiri and Tiwari (in press) have also proved the asymptotic optimality of their EB estimators under average squared error loss without requiring the n_i to be all equal. But, as pointed out earlier, they have not used any auxiliary information.

5.2 Bayes Estimation of a Quadratic Form when Ratios of Variance Components Known

We will consider the set up described in Section 3.2. We are interested in estimating a quadratic form $Y^T F Y$ where F ($N_T \times N_T$) is a known symmetric matrix. Writing

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \text{ where } F_{11} \text{ } (n_T \times n_T), F_{12} \text{ } (n_T \times (N_T - n_T))$$

and $F_{22} = ((N_T - n_T) \times (N_T - n_T))$ we can break up $Y^T F Y$ into three parts given by

$$\underline{Y}^T \underline{F} \underline{Y} = \underline{Y}^{(1)T} \underline{F}_{11} \underline{Y}^{(1)} + 2 \underline{Y}^{(1)T} \underline{F}_{12} \underline{Y}^{(2)} + \underline{Y}^{(2)T} \underline{F}_{22} \underline{Y}^{(2)}. \quad (5.2.1)$$

Under squared error loss, from (5.2.1) the Bayes estimator of $\underline{Y}^T \underline{F} \underline{Y}$ is given by

$$\begin{aligned} e_B(\underline{Y}^{(1)}) &= E(\underline{Y}^T \underline{F} \underline{Y} | \underline{Y}^{(1)}) \\ &= \underline{Y}^{(1)T} \underline{F}_{11} \underline{Y}^{(1)} + 2 \underline{Y}^{(1)T} \underline{F}_{12} E(\underline{Y}^{(2)} | \underline{Y}^{(1)}) \\ &\quad + E(\underline{Y}^{(2)T} \underline{F}_{22} \underline{Y}^{(2)} | \underline{Y}^{(1)}). \end{aligned} \quad (5.2.2)$$

Now from Theorem 3.2.1 we have

$$E(\underline{Y}^{(2)} | \underline{Y}^{(1)}) = \underline{M} \underline{Y}^{(1)} \quad (5.2.3)$$

and

$$\begin{aligned} V(\underline{Y}^{(2)} | \underline{Y}^{(1)}) &= (n_T + g_0 - p - 2)^{-1} (a_0 + \underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)}) \underline{G} \end{aligned} \quad (5.2.4)$$

where $\underline{K}$, $\underline{M}$ and $\underline{G}$ are given by (2.3.4) - (2.3.6). Using (5.2.3) and (5.2.4) in (5.2.2), we have

$$\begin{aligned} e_B(\underline{Y}^{(1)}) &= \underline{Y}^{(1)T} \underline{F}_{11} \underline{Y}^{(1)} + 2 \underline{Y}^{(1)T} \underline{F}_{12} \underline{M} \underline{Y}^{(1)} \\ &\quad + (\underline{M} \underline{Y}^{(1)})^T \underline{F}_{22} (\underline{M} \underline{Y}^{(1)}) + (n_T - g_0 - p - 2)^{-1} \\ &\quad \times (a_0 + \underline{Y}^{(1)T} \underline{K} \underline{Y}^{(1)}) \text{tr}(\underline{F}_{22} \underline{G}). \end{aligned} \quad (5.2.5)$$

From (5.2.5), we will derive the HB estimator of $\rho_i = (N_i - 1)^{-1} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2$, $i = 1, \dots, m$. Recall that $\underline{Y}^{(1)} = {}_{1 \leq i \leq m} \text{col}(\underline{Y}_i^{(1)})$, $\underline{Y}^{(2)} = {}_{1 \leq i \leq m} \text{col}(\underline{Y}_i^{(2)})$ and $\underline{Y}_i^T = (\underline{Y}_i^{(1)T}, \underline{Y}_i^{(2)T})$, $i = 1, \dots, m$. Also, let $\underline{F}_i = \begin{bmatrix} \underline{F}_{i11} & \underline{F}_{i12} \\ \underline{F}_{i21} & \underline{F}_{i22} \end{bmatrix}$, where $\underline{F}_{i11} = (N_i - 1)^{-1}(\underline{I}_{n_i} - N_i^{-1}\underline{J}_{n_i})$, $\underline{F}_{i12} = -(N_i - 1)^{-1}N_i^{-1}\underline{J}_{n_i, N_i - n_i} = \underline{F}_{i21}^T$ and $\underline{F}_{i22} = (N_i - 1)^{-1}(\underline{I}_{N_i - n_i} - N_i^{-1}\underline{J}_{N_i - n_i})$. It is easy to see that with these notations $\rho_i = \underline{Y}_i^T \underline{F}_i \underline{Y}_i$. Denote $E(\underline{Y}_i^{(2)} | \underline{Y}_i^{(1)}) = (\underline{M}\underline{Y}^{(1)})_i$ and $V(\underline{Y}_i^{(2)} | \underline{Y}_i^{(1)}) = \underline{G}_i$. Then from (5.2.5), we can write the Bayes estimator $\hat{\rho}_{iB}$ of ρ_i as

$$\begin{aligned} \hat{\rho}_{iB} = & \underline{Y}_i^{(1)T} \underline{F}_{i11} \underline{Y}_i^{(1)} + 2\underline{Y}_i^{(1)T} \underline{F}_{i12} (\underline{M}\underline{Y}^{(1)})_i \\ & + (\underline{M}\underline{Y}^{(1)})_i^T \underline{F}_{i22} (\underline{M}\underline{Y}^{(1)})_i + \text{tr}(\underline{F}_{i22} \underline{G}_i). \end{aligned} \quad (5.2.6)$$

We will use (5.2.6) in the following sections to find the Bayes predictor of $\underline{\rho}$ in the special case of nested error regression model.

5.3 Asymptotic Optimality in Nested Error Regression Model for Known λ

We will consider the model of Section 4.4 with $\lambda = \lambda_0$ (known), where the n_i are possibly unequal. For our convenience we will rewrite the model explicitly below.

I. Conditional on $\Theta = \theta$, $B = b$ and $R = r$,

$$Y_i \sim N(\theta_i 1_{N_i}, r^{-1} I_{N_i}), \quad i = 1, \dots, m \text{ independently};$$

II. conditional on $B = b$ and $R = r$, $\Theta \sim N(X_* b, (\lambda_0 r)^{-1} I_m)$

where $X_* = \underset{1 \leq i \leq m}{\text{col}} (x_i^T)$ is assumed to have full column rank $p < m$.

III. B and R are mutually independently distributed

with $B \sim \text{uniform}(R^p)$ and $R \sim \text{gamma}(\frac{1}{2}a_0, \frac{1}{2}g_0)$ where

$a_0 \geq 0$ and g_0 (real) such that $n_T + g_0 - p > 2$.

We can write I and II as a linear model

$$Y_{ij} = x_i^T b + v_i + e_{ij}, \quad (5.3.1)$$

$j = 1, \dots, N_i$; $i = 1, \dots, m$. All the random variables v_i and e_{ij} are mutually independent with $e_{ij} \sim N(0, r^{-1})$ and $v_i \sim N(0, (\lambda_0 r)^{-1})$, $j = 1, \dots, N_i$; $i = 1, \dots, m$.

From (5.3.1), we can write down X , Z , Ψ and D as appeared in (2.2.2). Using this, after some simplifications, it follows that after writing $u_i = \lambda_0 / (n_i + \lambda_0)$, $i = 1, \dots, m$, and $\hat{b} = \left[\sum_{\alpha=1}^m (1 - u_\alpha) x_\alpha x_\alpha^T \right]^{-1} \sum_{\alpha=1}^m (1 - u_\alpha) x_\alpha \bar{Y}_{\alpha(s)}$,

$$\begin{aligned} (MY^{(1)})_i &= E(Y_i^{(2)} | Y^{(1)}) \\ &= (n_i + \lambda_0)^{-1} (n_i \bar{Y}_{i(s)} + \lambda_0 x_i^T \hat{b}) 1_{N_i - n_i} \\ &= \left[(1 - u_i) \bar{Y}_{i(s)} + u_i x_i^T \hat{b} \right] 1_{N_i - n_i}, \end{aligned} \quad (5.3.2)$$

$$\begin{aligned}
G_i &= v(Y_i^{(2)} | Y^{(1)}) \\
&= (n_T + g_0 - p - 2)^{-1} (a_0 + Y^{(1)T} K Y^{(1)}) \left\{ I_{N_i - n_i} \right. \\
&\quad \left. + \lambda_0^{-1} u_i J_{N_i - n_i} + \lambda_0^{-1} u_i^2 x_i^T \left(\sum_{\alpha=1}^m (1 - u_\alpha) x_\alpha x_\alpha^T \right)^{-1} x_i J_{N_i - n_i} \right\}.
\end{aligned}
\tag{5.3.3}$$

Using (5.3.2) and (5.3.3), we have from (5.2.6), after some simplifications,

$$\begin{aligned}
\hat{\rho}_{iB} &= (N_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i(s)})^2 \\
&\quad + (N_i - 1)^{-1} n_i f_i u_i^2 (\bar{Y}_{i(s)} - x_i^T \hat{b})^2 + (N_i - 1)^{-1} \\
&\quad \times f_i \left[N_i - u_i + u_i (1 - u_i) x_i^T \left(\sum_{\alpha=1}^m a_\alpha a_\alpha^T \right)^{-1} x_i \right] \\
&\quad \times (a_0 + Y^{(1)T} K Y^{(1)}) / (n_T + g_0 - p - 2)
\end{aligned}
\tag{5.3.4}$$

where $a_i = (1 - u_i)^{\frac{1}{2}} x_i$, $i = 1, \dots, m$.

We will now show that $\hat{\rho}_{iB}$, $i = 1, \dots, m$, is asymptotically optimal under the loss (5.1.1) and under the subjective prior π_0 which specifies that $Y_i \sim N\left((x_i^T b_0) 1_{N_i}, r_0^{-1} (I_{N_i} + \lambda_0^{-1} J_{N_i})\right)$, $i = 1, \dots, m$ independently. To this end, we will have to find the subjective Bayes estimator $\hat{\rho}_{iSB}$ of ρ_i . Under the loss (5.1.1), for $i = 1, \dots, m$, after some simplifications

$$\begin{aligned}
\hat{\rho}_{iSB} &= E_{\pi_0}(\rho_i | Y^{(1)}) \\
&= (N_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i(s)})^2 + n_i f_i (N_i - 1)^{-1} u_i^2 \\
&\quad \times (\bar{Y}_{i(s)} - x_i^T b_0)^2 + (N_i - 1)^{-1} r_0^{-1} f_i (N_i - u_i)
\end{aligned} \tag{5.3.5}$$

where E_{π_0} denotes the expectation taken under the prior distribution π_0 .

Let $\hat{\rho}_B = (\hat{\rho}_{1B}, \dots, \hat{\rho}_{mB})^T$ and $\hat{\rho}_{SB} = (\hat{\rho}_{1SB}, \dots, \hat{\rho}_{mSB})^T$.

Using the notations of Chapter Four, to prove the A.O. for $\hat{\rho}_B$, we need to show

$$r_{Q_m}(\pi_0, \hat{\rho}_B) - r_{Q_m}(\pi_0, \hat{\rho}_{SB}) \rightarrow 0 \text{ as } m \rightarrow \infty.$$

We establish it in the following theorem assuming the condition (4.2.2.1).

Theorem 5.3.1. Assume that the condition (4.2.2.1) holds. Then for the prior π_0 given above, the model (5.3.1) and the loss (5.1.1), $\hat{\rho}_B$, the HB predictor of ρ is asymptotically optimal.

Proof of Theorem 5.3.1. Using standard Bayesian calculations, we get

$$\begin{aligned}
r_{Q_m}(\pi_0, \hat{\rho}_B) - r_{Q_m}(\pi_0, \hat{\rho}_{SB}) \\
= m^{-1} E_{\pi_0} \left[(\hat{\rho}_B - \hat{\rho}_{SB})^T Q_m (\hat{\rho}_B - \hat{\rho}_{SB}) \right] = a_m \text{ (say)}.
\end{aligned}$$

Since $a_m \geq 0 \forall m$, it is enough to prove $\overline{\lim}_{m \rightarrow \infty} a_m = 0$. By Lemma 4.2.2.1,

$$a_m \leq \text{ch}_L(Q_m) m^{-1} \sum_{i=1}^m E_{\pi_0} (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2$$

and hence by condition (4.2.2.1)

$$\begin{aligned} \overline{\lim}_{m \rightarrow \infty} a_m &\leq \overline{\lim}_{m \rightarrow \infty} \text{ch}_L(Q_m) \overline{\lim}_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2 \\ &\leq \Gamma \overline{\lim}_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2. \end{aligned}$$

So it is enough to show that

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2 = 0. \quad (5.3.6)$$

Now,

$$\begin{aligned} &(N_i - 1)(\hat{\rho}_{iB} - \hat{\rho}_{iSB}) \\ &= n_i f_i u_i^2 \left\{ (\bar{Y}_{i(s)} - \bar{x}_i^T \hat{b})^2 - (\bar{Y}_{i(s)} - \bar{x}_i^T b_0)^2 \right\} \\ &\quad + f_i (N_i - u_i) \left\{ (a_0 + Y^{(1)T} K Y^{(1)}) \right. \\ &\quad \times (n_T + g_0 - p - 2)^{-1} - r_0^{-1} \left. \right\} \\ &\quad + f_i u_i (1 - u_i) \bar{x}_i^T \left(\sum_{\alpha=1}^m \bar{a}_\alpha \bar{a}_\alpha^T \right)^{-1} \bar{x}_i (a_0 + Y^{(1)T} K Y^{(1)}) \\ &\quad \times (n_T + g_0 - p - 2)^{-1}, \end{aligned} \quad (5.3.7)$$

and

$$\begin{aligned}
(\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 - (\bar{Y}_{i(s)} - \mathbf{x}_i^T \mathbf{b}_0)^2 &= t_i \text{ (say)} \\
&= - \left[\mathbf{x}_i^T (\hat{\mathbf{b}} - \mathbf{b}_0) \right]^2 - 2\mathbf{x}_i^T (\hat{\mathbf{b}} - \mathbf{b}_0) (\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}}).
\end{aligned}$$

Note that under π_0 , $\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}}$ and $\hat{\mathbf{b}}$ are independently distributed with $\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}} \sim N\left(0, r_0^{-1}(n_i^{-1}u_i^{-1} - \lambda_0^{-1}\mathbf{x}_i^T \times (\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T)^{-1} \mathbf{x}_i)\right)$ and $\hat{\mathbf{b}} \sim N(\mathbf{b}_0, (\lambda_0 r_0)^{-1}(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T)^{-1})$. Then after some simplifications

$$E_{\pi_0}(t_i) = -(\lambda_0 r_0)^{-1} \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i \quad (5.3.8)$$

and

$$\begin{aligned}
V_{\pi_0}(t_i) &= 2(r_0 \lambda_0)^{-2} \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i \\
&\quad \times \left[2u_i^{-1} - \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i \right] \\
&\leq 4(r_0 \lambda_0)^{-2} \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i u_i^{-1}. \quad (5.3.9)
\end{aligned}$$

From (5.3.7) and (5.3.8),

$$\begin{aligned}
(N_i - 1)(\hat{\rho}_{iB} - \hat{\rho}_{iSB}) \\
&= f_i \left[n_i u_i^2 (t_i - E_{\pi_0}(t_i)) \right. \\
&\quad \left. + h_i \left((\mathbf{a}_0 + \mathbf{Y}^{(1)T} \mathbf{K} \mathbf{Y}^{(1)}) / (n_T + g_0 - p - 2) - r_0^{-1} \right) \right]
\end{aligned}$$

where $h_i = (N_i - u_i) + u_i(1 - u_i) \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i$. Note that

$(1 - u_i) \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i = \mathbf{a}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{a}_i$ and the matrix $\mathbf{P} = \left(\mathbf{a}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{a}_j \right)$ is symmetric idempotent. Therefore $\mathbf{a}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{a}_i \leq 1$. Using this we get $h_i(N_i - 1)^{-1} \leq (N_i - 1)^{-1} [N_i - u_i + u_i] \leq 2$ since we will assume that $N_i \geq n_i + 1 \geq 3$. From all these observations, it follows that

$$\begin{aligned} & (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2 \\ & \leq 2 \left[n_i^2 u_i^4 (t_i - E\pi_0(t_i))^2 \right. \\ & \quad \left. + 4 \left((a_0 + \mathbf{Y}^{(1)T} \mathbf{K} \mathbf{Y}^{(1)}) / (n_T + g_0 - p - 2) - r_0^{-1} \right)^2 \right]. \end{aligned}$$

Hence

$$\begin{aligned} m^{-1} \sum_{i=1}^m (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2 & \leq 2m^{-1} \sum_{i=1}^m n_i^2 u_i^4 (t_i - E\pi_0(t_i))^2 \\ & \quad + 8 \left[(a_0 + \mathbf{Y}^{(1)T} \mathbf{K} \mathbf{Y}^{(1)}) / (n_T + g_0 - p - 2) - r_0^{-1} \right]^2. \end{aligned} \quad (5.3.10)$$

From (5.3.9) and (5.3.10) we have

$$\begin{aligned} & m^{-1} \sum_{i=1}^m E\pi_0 (\hat{\rho}_{iB} - \hat{\rho}_{iSB})^2 \\ & \leq 8(r_0 \lambda_0)^{-2} m^{-1} \sum_{i=1}^m u_i^3 n_i^2 \mathbf{x}_i^T \left(\sum_{\alpha=1}^m \mathbf{a}_\alpha \mathbf{a}_\alpha^T \right)^{-1} \mathbf{x}_i \\ & \quad + 8E\pi_0 \left[(a_0 + \mathbf{Y}^{(1)T} \mathbf{K} \mathbf{Y}^{(1)}) / (n_T + g_0 - p - 2) - r_0^{-1} \right]^2. \end{aligned} \quad (5.3.11)$$

Now since $u_i = \lambda_0 / (\lambda_0 + n_i)$ and $a_i = (n_i / (\lambda_0 + n_i))^{\frac{1}{2}} x_i$,

$$\begin{aligned}
 & 8(r_0 \lambda_0)^{-2m-1} \sum_{i=1}^m u_i^3 n_i^2 x_i^T \left(\sum_{\alpha=1}^m a_\alpha a_\alpha^T \right)^{-1} x_i \\
 &= 8r_0^{-2m-1} \sum_{i=1}^m u_i (1 - u_i) a_i^T \left(\sum_{\alpha=1}^m a_\alpha a_\alpha^T \right)^{-1} a_i \\
 &\leq 2r_0^{-2m-1} \sum_{i=1}^m a_i^T \left(\sum_{\alpha=1}^m a_\alpha a_\alpha^T \right)^{-1} a_i \\
 &= 2r_0^{-2m-1} \text{tr} \left[\left(\sum_{\alpha=1}^m a_\alpha a_\alpha^T \right)^{-1} \left(\sum_{i=1}^m a_i a_i^T \right) \right] \\
 &= 2r_0^{-2m-1} p. \tag{5.3.12}
 \end{aligned}$$

Recall that $\Sigma_{11} = \bigoplus_{i=1}^m [I_{n_i} + \lambda_0^{-1} J_{n_i}]$, $X^{(1)} = \text{col}_{1 \leq i \leq m} (1_{n_i} x_i^T)$ and

$K = \Sigma_{11}^{-1} - \Sigma_{11}^{-1} X^{(1)} (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} X^{(1)T} \Sigma_{11}^{-1}$. Since under π_0 ,

$Y^{(1)} \sim N(X^{(1)} b_0, r_0^{-1} \Sigma_{11})$, it follows from Rao (1973, see Result (vii) on p. 188) that $Y^{(1)T} K Y^{(1)} \sim r_0^{-1} \chi_{n_T - p}^2$. Then

it is routine to check that

$$\lim_{m \rightarrow \infty} E \pi_0 \left[\left(a_0 + Y^{(1)T} K Y^{(1)} \right) / (n_T + g_0 - p - 2) - r_0^{-1} \right]^2 = 0 \tag{5.3.13}$$

since $n_T \rightarrow \infty$ as $m \rightarrow \infty$.

Combining (5.3.11) - (5.3.13), (5.3.6) follows. This completes the proof of Theorem 5.3.1.

Let $\hat{\rho}_U = (s_1^2, \dots, s_m^2)^T$ where $s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i(s))^2$ assuming $n_i \geq 2$, $i = 1, \dots, m$. We will show now in the next

theorem that under some conditions, this estimator is not asymptotically optimal.

Theorem 5.3.2. Assume the conditions (4.2.2.9) - (4.2.2.11) hold. Then for the prior π_0 given above, the model (5.3.1) and the loss (5.1.1), $\hat{\rho}_U$, the traditional estimator of ρ , is asymptotically nonoptimal.

Proof of Theorem 5.3.2. Let $d_m = r_{Q_m}(\pi_0, \hat{\rho}_U) - r_{Q_m}(\pi_0, \hat{\rho}_{SB})$. Then

$$\begin{aligned} d_m &= m^{-1} E_{\pi_0} (\hat{\rho}_U - \hat{\rho}_{SB})^T Q_m (\hat{\rho}_U - \hat{\rho}_{SB}) \\ &\geq \text{ch}_S(Q_m) m^{-1} \sum_{i=1}^m E_{\pi_0} (s_i^2 - \hat{\rho}_{iSB})^2, \quad \text{by Lemma 4.2.2.1} \end{aligned}$$

and hence

$$\lim_{m \rightarrow \infty} d_m \geq \omega \lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} (s_i^2 - \hat{\rho}_{iSB})^2 \quad \text{by (4.2.2.11).}$$

It is enough to prove that

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} (s_i^2 - \hat{\rho}_{iSB})^2 > 0.$$

After some simplifications,

$$\begin{aligned} \hat{\rho}_{iSB} - s_i^2 &= [(n_i - 1)(N_i - 1)^{-1} - 1](s_i^2 - r_0^{-1}) \\ &\quad + n_i(N_i - 1)^{-1} f_{i u_i}^2 \left[(\bar{Y}_{i(s)} - x_i^T b_0)^2 - r_0^{-1} n_i^{-1} u_i^{-1} \right]. \end{aligned}$$

Under π_0 , s_i^2 and $\bar{Y}_{i(s)}$ are independently distributed for $i = 1, \dots, m$. Moreover $E_{\pi_0}(s_i^2) = r_0^{-1}$. Then

$$\begin{aligned}
& E_{\pi_0} [s_i^2 - \hat{\rho}_{iSB}]^2 \\
&= [(N_i - n_i)/(N_i - 1)]^2 V_{\pi_0}(s_i^2) + n_i^2 (N_i - 1)^{-2} f_i^4 u_i^4 \\
&\quad \times E_{\pi_0} \left[(\bar{Y}_{i(s)} - \mathbf{x}_i^T \mathbf{b}_0)^2 - r_0^{-1} n_i^{-1} u_i^{-1} \right]^2 \\
&\geq 2r_0^{-2} f_i^2 / (n_i - 1) \\
&\geq 2r_0^{-2} (1 - n_i / (n_i + 1))^2 / (n_i - 1) \quad \text{by (4.2.2.10)} \\
&\geq 2r_0^{-2} (K + 1)^{-2} (K - 1)^{-1} \quad \text{by (4.2.2.9)}.
\end{aligned}$$

Hence

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E_{\pi_0} [s_i^2 - \hat{\rho}_{iSB}]^2 \geq 2r_0^{-2} (K + 1)^{-2} (K - 1)^{-1} > 0.$$

The proof of Theorem 5.3.2 is complete.

5.4 Asymptotic Optimality in Nested Error Regression Model with Unknown Variance Components

In the previous section, we assumed the ratio of variance components was known. Here we will dispense with this assumption and use the hierarchical model described in Section 4.4. This is an extension of the model in Section 5.3 where λ ($= \lambda_0$) is assumed to be known. However, unlike in Section 5.3, we will assume all the n_i are equal to n .

We will first derive the HB predictor $\hat{\rho}_{HB}$ of ρ . Note that the HB predictor $\hat{\rho}_{iHB}$ of ρ_i is given by

$$\begin{aligned}\hat{\rho}_{iHB} &= E[\rho_i | Y^{(1)}] \\ &= E[E(\rho_i | \Lambda, Y^{(1)}) | Y^{(1)}].\end{aligned}\quad (5.4.1)$$

To facilitate the derivation of $E(\rho_i | \Lambda, Y^{(1)})$ in (5.4.1) we will write down the expressions of the first two moments of the distribution of $Y_i^{(2)}$ given $Y^{(1)}$ and Λ . From III of Section 4.4, it easily follows that

IV. conditional on $\Lambda = \lambda$, B and R are mutually independently distributed with $B \sim \text{uniform}(R^p)$ and $R \sim \text{gamma}(\frac{1}{2}a_0, \frac{1}{2}(g_0 - 2\alpha))$.

Write $U = \Lambda/(n + \Lambda)$ and $u = \lambda/(n + \lambda)$. From I, II and IV, we have as in (5.3.2) and (5.3.3) with appropriate modifications that

$$E(Y_i^{(2)} | \Lambda, Y^{(1)}) = [(1 - U)\bar{Y}_{i(s)} + U\mathbf{x}_i^T \hat{\mathbf{b}}] 1_{N_i - n}, \quad (5.4.2)$$

$$\begin{aligned}V(Y_i^{(2)} | \Lambda, Y^{(1)}) &= (n_T + g_0 - 2\alpha - p - 2)^{-1} (a_0 + Y^{(1)T} K Y^{(1)}) \\ &\quad \times \left\{ I_{N_i - n} + \Lambda^{-1} U J_{N_i - n} + n^{-1} U \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i J_{N_i - n} \right\}.\end{aligned}\quad (5.4.3)$$

In this balanced situation, $n_T = nm$, $Y^{(1)T} K Y^{(1)} = S + \text{USSR}$ where $S = \sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i(s)})^2$ and $\text{SSR} = n \bar{Y}_{(s)}^T (I_m - P_{X_*}) \bar{Y}_{(s)}$ and $\hat{\mathbf{b}} = (X_*^T X_*)^{-1} X_*^T \bar{Y}_{(s)}$. Note that $\hat{\mathbf{b}}$ does not depend on Λ .

As in Section 5.3, we get from (5.4.1) ~ (5.4.3) that

$$\begin{aligned}
 E(\rho_i | \Lambda, Y^{(1)}) &= (N_i - 1)^{-1} \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i(s)})^2 + (N_i - 1)^{-1} n f_i U^2 \\
 &\quad \times (\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 + (N_i - 1)^{-1} f_i \\
 &\quad \times \left[N_i - U + U(1 - U) \mathbf{x}_i^T \left(\sum_{k=1}^m (1 - U) \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right] \\
 &\quad \times (a_0 + S + \text{USSR}) / (nm + g_0 - 2\alpha - p - 2).
 \end{aligned}$$

Finally, we have

$$\begin{aligned}
 \hat{\rho}_{iHB} &= (N_i - 1)^{-1} \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i(s)})^2 \\
 &\quad + (N_i - 1)^{-1} n f_i (\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 E_\alpha(U^2 | \bar{Y}_{i(s)}, S) \\
 &\quad + (N_i - 1)^{-1} f_i E_\alpha \left[\left\{ N_i - U \left(1 - \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right) \right\} \right. \\
 &\quad \times \left. (a_0 + S + \text{USSR}) \middle| \bar{Y}_{i(s)}, S \right] / (nm + g_0 - 2\alpha - p - 2)
 \end{aligned} \tag{5.4.4}$$

where $E_\alpha(\cdot)$ is defined in (4.4.5).

Now we will show the risk difference $r_{Q_m}(\pi_0, \hat{\rho}_{HB}) - r_{Q_m}(\pi_0, \hat{\rho}_{SB}) = a_m$ (say) goes to zero as $m \rightarrow \infty$ to show that the HB estimator is asymptotically optimal. The prior π_0 is as given in Section 5.3 (also in Section 4.4). We will

state and prove the following theorem. Note that α appearing in the theorem is a prior parameter.

Theorem 5.4.1. Assume the condition (4.2.2.1) holds. Then for the prior π_0 given above and for $\alpha \geq 1$, $\hat{\rho}_{HB}$ is asymptotically optimal for the balanced nested error regression model.

Proof of Theorem 5.4.1. As in Theorem 4.4.1, let E^* denote the expectation taken w.r.t. the joint distribution of $\bar{Y}_{(s)}$ and S specified by the prior π_0 . Then following the proof of Theorem 5.3.1 it is enough to show that

$$m^{-1} \sum_{i=1}^m E^* [\hat{\rho}_{iHB} - \hat{\rho}_{iSB}]^2 \rightarrow 0 \text{ as } m \rightarrow \infty.$$

Let $E_\alpha(U | \bar{Y}_{(s)}, S) = W_m(\bar{Y}_{(s)}, S; \alpha) \equiv W_m$, $E_\alpha(U^2 | \bar{Y}_{(s)}, S) = V_m(\bar{Y}_{(s)}, S; \alpha) \equiv V_m$ and $u_0 = \lambda_0 / (n + \lambda_0)$. Then

$$\begin{aligned} & (\hat{\rho}_{iHB} - \hat{\rho}_{iSB}) \\ &= (N_i - 1)^{-1} f_{in} \left[(\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 V_m - (\bar{Y}_{i(s)} - \mathbf{x}_i^T \mathbf{b}_0)^2 u_0^2 \right] \\ &+ (N_i - 1)^{-1} f_i E_\alpha \left[\left\{ N_i - U \left(1 - \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right) \right\} \right. \\ &\quad \left. \times (a_0 + S + USSR) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} (N_i - u_0) \right] \bar{Y}_{i(s)}, S \Big] \\ &= f_i [t_{1i} + t_{2i} + t_{3i} + t_{4i}] \end{aligned} \tag{5.4.5}$$

where

$$t_{1i} = (N_i - 1)^{-1} n (\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 (v_m - u_0^2), \quad (5.4.6)$$

$$t_{2i} = (N_i - 1)^{-1} n u_0^2 \left\{ (\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 - (\bar{Y}_{i(s)} - \mathbf{x}_i^T \mathbf{b}_0)^2 \right. \\ \left. + n^{-1} r_0^{-1} u_0^{-1} \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right\}, \quad (5.4.7)$$

$$t_{3i} = (1 - N_i^{-1})^{-1} \{ (a_0 + S + SSRW_m) \\ \div (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} \}, \quad (5.4.8)$$

and

$$t_{4i} = -(N_i - 1)^{-1} \left\{ 1 - \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right\} \\ \times E_\alpha \left[U(a_0 + S + USSR) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} u_0 \right] \bar{Y}_{i(s)}, S \Big]. \quad (5.4.9)$$

Then

$$(\hat{\rho}_{iHB} - \hat{\rho}_{iSB})^2 \leq 4 \sum_{j=1}^4 t_{ji}^2 \\ \text{since } f_i \leq 1 \text{ and } \left(\sum_{j=1}^4 t_{ji} \right)^2 \leq 4 \sum_{j=1}^4 t_{ji}^2.$$

Consequently,

$$m^{-1} \sum_{i=1}^m (\hat{\rho}_{iHB} - \hat{\rho}_{iSB})^2 \leq 4 \sum_{j=1}^4 m^{-1} \sum_{i=1}^m t_{ji}^2. \quad (5.4.10)$$

As in Lemma 4.4.2, we can show that $v_m \xrightarrow{a.s.} u_0^2$ under π_0 . Using this, we can show that $m^{-1} \sum_{i=1}^m t_{1i}^2 \xrightarrow{a.s.} 0$. Also $m^{-1} \sum_{i=1}^m t_{1i}^2$ is uniformly integrable. So we have

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E^*(t_{1i}^2) = 0. \quad (5.4.11)$$

Now since

$$E^*(t_{2i}^2) = (N_i - 1)^{-2} n^2 u_0^2 v \pi_0 \left[(\bar{Y}_{i(s)} - \mathbf{x}_i^T \hat{\mathbf{b}})^2 - (\bar{Y}_{i(s)} - \mathbf{x}_i^T \mathbf{b}_0)^2 \right],$$

it follows from a calculation in the proof of Theorem 5.3.1 that

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E^*(t_{2i}^2) = 0. \quad (5.4.12)$$

Since $N_i > 2$ for all i , we have

$$\begin{aligned} m^{-1} \sum_{i=1}^m t_{3i}^2 &\leq 2.25 \left\{ (a_0 + S + SSRW_m) \right. \\ &\quad \left. \div (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} \right\}^2. \end{aligned} \quad (5.4.13)$$

Recall that $W_m \xrightarrow{a.s.} u_0$. Also it can be shown that under π_0 , $S/(m(n-1)) \xrightarrow{a.s.} r_0^{-1}$ and $SSR/(m-p) \xrightarrow{a.s.} r_0^{-1} u_0^{-1}$, by using Markov's inequality and Borel-Cantelli lemma as in Section 4.3. From these, it follows that

$$\left[(a_0 + S + SSRW_m) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} \right]^2 \xrightarrow{a.s.} 0.$$

Also it can be shown uniformly integrable. Hence

$$E^* \left[(a_0 + S + SSRW_m) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} \right]^2 \rightarrow 0 \text{ as } m \rightarrow \infty. \text{ Therefore, from (5.4.13)}$$

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E^*(t_{3i}^2) = 0. \quad (5.4.14)$$

Note that since $0 \leq \mathbf{x}_i^T \left(\sum_{k=1}^m \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \leq 1$, we have $t_{4i}^2 \leq \left\{ (W_m(a_0 + S) + SSRV_m) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1} u_0 \right\}^2$.

We can also show as before

$$\left[(W_m(a_0 + S) + SSRV_m) / (nm + g_0 - 2\alpha - p - 2) - r_0^{-1}u_0 \right]^2 \xrightarrow{a.s.} 0$$

and uniformly integrable. From this, we have

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E^*(t_{4i}^2) = 0. \quad (5.4.15)$$

Combining (5.4.10) - (5.4.15) we get

$$\lim_{m \rightarrow \infty} m^{-1} \sum_{i=1}^m E^*[\hat{\rho}_{iHB} - \hat{\rho}_{iSB}]^2 = 0$$

and the theorem follows.

CHAPTER SIX SUMMARY AND FUTURE RESEARCH

6.1 Summary

In this dissertation, a unified model-based HB prediction theory is developed for small area estimation as well as for animal breeding and comparative experiments. On the basis of the analysis of two data sets in Chapter Two, it is apparent that the HB analysis is a viable alternative to the existing frequentist methods of inference. Especially, for complex models, often we do not have closed form expressions, or suitable approximations, of the MSEs of the EBLUPs or the EB predictors. On the other hand, the Bayesian procedures provided in this dissertation can be used routinely given the present state of computing facilities.

In the special case of known ratios of variance components, the HB predictor was shown to be BLUP for a vector of linear functions of the finite population observation vector or the vector of effects, without any distributional assumption. Moreover, for a suitable class of elliptically symmetric distributions, it was shown (a) to be the BUP, (b) to universally or stochastically

dominate all linear unbiased predictors, and (c) to be the best equivariant predictor under suitable groups of transformations. So if one has a fairly good idea of the approximate values of the variance components based on past data, one may incorporate that information in the prior and expect that the proposed HB predictor in the general case will be approximately optimal. This was supported by the asymptotic optimality property we established for a few specific models. Also, we proved the asymptotic optimality of an HB predictor of the finite population variance vector in an important special case.

6.2 Future Research

We have proposed an HB model which is applicable when we have a single characteristic for each unit in the population. In survey sampling and animal breeding experiments, it is quite common to have more than one characteristic for each unit and characteristics are correlated within a unit. A useful research will be to provide a suitable multivariate extension of the proposed model to include this type of problems. Another important model which is not considered is the longitudinal data problem. Here for each individual unit we measure one or more characteristics over time. An HB model, under such circumstances, involves generalization of the present model

to take into account both the within unit and between unit variation.

The present dissertation did not address the robustness issue except for the study of universal and stochastic domination which in some sense provides robustness for the bowl-shaped loss functions. One important issue is the prior robustness. For the vector of fixed effects, we have used uniform prior which can be viewed as a diffused multivariate normal prior. The performance of the proposed HB predictor for other priors whose tails are heavier than the normal (e.g., multivariate t) is a topic for future study.

In the preceding two chapters, we have studied the asymptotic optimality of the HB predictors for special models with two variance components when both variance components are unknown. It might be worth exploring whether the techniques used there can be generalized to problems with more than two variance components.

Finally, in this dissertation, we assume no nonsampling error, measurement error, bias or nonresponse so that once a sample is drawn, the value of the characteristic is known for sure. So in predicting finite population means or variances, we did not change the sampled vector which corresponds to the seen part of the characteristic vector. Estimation of the finite population

means or variances in presence of response bias or measurement error can also be explored.

APPENDICES

APPENDIX A
PROOF OF THEOREM 2.3.1

Under the assumptions of the Theorem 2.3.1, the joint pdf of $\underline{Y}$, $\underline{B}$, R and $\underline{\Lambda}$ is given by

$$\begin{aligned}
 f(\underline{y}, \underline{b}, r, \underline{\lambda}) & \propto r^{\frac{1}{2}N} |\underline{\Sigma}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} r (\underline{y} - \underline{X}\underline{b})^T \underline{\Sigma}^{-1} (\underline{y} - \underline{X}\underline{b}) \right] \\
 & \times \exp \left(-\frac{1}{2} a_0 r \right) r^{\frac{1}{2}g_0 - 1} \exp \left(-\frac{1}{2} r \sum_{i=1}^t a_i \lambda_i \right) \prod_{i=1}^t (\lambda_i r)^{\frac{1}{2}g_i - 1} r^t \\
 & = |\underline{\Sigma}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} r \left\{ (\underline{y} - \underline{X}\underline{b})^T \underline{\Sigma}^{-1} (\underline{y} - \underline{X}\underline{b}) + a_0 + \sum_{i=1}^t a_i \lambda_i \right\} \right] \\
 & \times r^{\frac{1}{2} \left(N_T + \sum_{i=0}^t g_i \right) - 1} \prod_{i=1}^t \lambda_i^{\frac{1}{2}g_i - 1}. \tag{A.1}
 \end{aligned}$$

Now,

$$(\underline{y} - \underline{X}\underline{b})^T \underline{\Sigma}^{-1} (\underline{y} - \underline{X}\underline{b}) = (\underline{b} - \hat{\underline{b}})^T \underline{X}^T \underline{\Sigma}^{-1} \underline{X} (\underline{b} - \hat{\underline{b}}) + \underline{y}^T \underline{Q} \underline{y}, \tag{A.2}$$

where $\hat{\underline{b}} = (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1} \underline{Y}$ and $\underline{Q} = \underline{\Sigma}^{-1} - \underline{\Sigma}^{-1} \underline{X} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1}$.

From (A.1) and (A.2), one gets, integrating $\underline{b}$ out, the joint pdf of $\underline{Y}$, R and $\underline{\Lambda}$ is given by

$$f(\underline{y}, r, \underline{\lambda}) \propto |\underline{\Sigma}|^{-\frac{1}{2}} |\underline{X}^T \underline{\Sigma}^{-1} \underline{X}|^{-\frac{1}{2}} r^{\frac{1}{2}(N_T + \sum_{i=1}^t g_i - p) - 1} \\ \times \exp \left[-\frac{1}{2} r \left(a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^T \underline{Q} \underline{y} \right) \right] \prod_{i=1}^t \lambda_i^{\frac{1}{2} g_i - 1}. \quad (\text{A.3})$$

Now, integrating w.r.t. r , one finds the joint pdf of $\underline{Y}$ and $\underline{\Lambda}$ is given by

$$f(\underline{y}, \underline{\lambda}) \propto |\underline{\Sigma}|^{-\frac{1}{2}} |\underline{X}^T \underline{\Sigma}^{-1} \underline{X}|^{-\frac{1}{2}} \\ \times \left(a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^T \underline{Q} \underline{y} \right)^{-\frac{1}{2}(N_T + \sum_{i=1}^t g_i - p)} \prod_{i=1}^t \lambda_i^{\frac{1}{2} g_i - 1}. \quad (\text{A.4})$$

Similarly starting with the joint pdf of $\underline{Y}^{(1)}$, $\underline{B}$, $\underline{R}$ and $\underline{\Lambda}$ one gets the joint pdf of $\underline{Y}^{(1)}$ and $\underline{\Lambda}$ is given by

$$h(\underline{y}^{(1)}, \underline{\lambda}) \propto |\underline{\Sigma}_{11}|^{-\frac{1}{2}} |\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)}|^{-\frac{1}{2}} \\ \times \left(a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right)^{-\frac{1}{2}(n_T + \sum_{i=1}^t g_i - p)} \\ \times \prod_{i=1}^t \lambda_i^{\frac{1}{2} g_i - 1}. \quad (\text{A.5})$$

So the conditional pdf of $\underline{Y}^{(2)}$ given $\underline{\Lambda} = \underline{\lambda}$ and $\underline{Y}^{(1)} = \underline{y}^{(1)}$ is given by

$$f(\underline{y}^{(2)} | \underline{\lambda}, \underline{y}^{(1)}) = \frac{f(\underline{y}, \underline{\lambda})}{h(\underline{y}^{(1)}, \underline{\lambda})}$$

$$\begin{aligned}
& \propto \left(|\Sigma|/|\Sigma_{11}| \right)^{-\frac{1}{2}} \left(|X^T \Sigma^{-1} X| / |X^{(1)T} \Sigma_{11}^{-1} X^{(1)}| \right)^{-\frac{1}{2}} \\
& \times \left(a_0 + \sum_{i=1}^t a_i \lambda_i + Y^{(1)T} K_Y^{(1)} \right)^{\frac{1}{2}(n_T + \sum_{i=0}^t g_i - p)} \\
& \times \left(a_0 + \sum_{i=1}^t a_i \lambda_i + Y^T Q_Y \right)^{-\frac{1}{2}(N_T + \sum_{i=0}^t g_i - p)}. \quad (A.6)
\end{aligned}$$

Now,

$$\begin{aligned}
X^T \Sigma^{-1} X &= X^{(1)T} \Sigma_{11}^{-1} X^{(1)} + (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T \\
&\times \Sigma_{22.1}^{-1} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})
\end{aligned}$$

gives

$$\begin{aligned}
|X^T \Sigma^{-1} X| &= \left| X^{(1)T} \Sigma_{11}^{-1} X^{(1)} + (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T \right. \\
&\quad \left. \times \Sigma_{22.1}^{-1} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) \right| \\
&= \left| \begin{array}{cc} X^{(1)T} \Sigma_{11}^{-1} X^{(1)} & -(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T \\ X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} & \Sigma_{22.1} \end{array} \right| \div |\Sigma_{22.1}| \\
&= \left| X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right| \Sigma_{22.1} + (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)})^T \\
&\quad \times (X^{(1)T} \Sigma_{11}^{-1} X^{(1)})^{-1} (X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}) \left| \right| \div |\Sigma_{22.1}|, \quad (A.7)
\end{aligned}$$

by using Exercise 2.4 on p. 32 of Rao (1973) twice. Again

using the same exercise, we have

$$|\Sigma| = |\Sigma_{11}| |\Sigma_{22.1}|. \quad (\text{A.8})$$

Using the definition of G from (2.3.6) and (A.7) and (A.8) we have

$$|X^T \Sigma^{-1} X| = |X^{(1)T} \Sigma_{11}^{-1} X^{(1)}| |G| \div (|\Sigma|/|\Sigma_{11}|)$$

i.e.,

$$\left(|X^T \Sigma^{-1} X| / |X^{(1)T} \Sigma_{11}^{-1} X^{(1)}| \right) (|\Sigma|/|\Sigma_{11}|) = |G|. \quad (\text{A.9})$$

Now using a result of Sallas and Harville (1981) we have

$$Q = \lim_{\epsilon \rightarrow \infty} (\Sigma + \epsilon X X^T)^{-1} = \lim_{\epsilon \rightarrow \infty} B^{-1}(\epsilon) \quad (\text{say}). \quad (\text{A.10})$$

We partition $B(\epsilon)$ into

$$B(\epsilon) = \begin{bmatrix} B_{11}(\epsilon) & B_{12}(\epsilon) \\ B_{21}(\epsilon) & B_{22}(\epsilon) \end{bmatrix}, \quad (\text{A.11})$$

where $B_{ij}(\epsilon) = \Sigma_{ij} + \epsilon X^{(i)} X^{(j)T}$ ($i, j = 1, 2$). Next using a standard formula for partitioned matrices (e.g., p. 46 of Searle, 1971),

$$\begin{aligned} \underline{y}^T \mathbb{B}^{-1}(\epsilon) \underline{y} &= \underline{y}^{(1)T} \mathbb{B}_{11}^{-1}(\epsilon) \underline{y}^{(1)} + \left(\underline{y}^{(2)} - \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) \underline{y}^{(1)} \right)^T \\ &\quad \times \mathbb{B}_{22.1}^{-1} \left(\underline{y}^{(2)} - \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) \underline{y}^{(1)} \right), \end{aligned} \quad (\text{A.12})$$

where $\mathbb{B}_{22.1}(\epsilon) = \mathbb{B}_{22}(\epsilon) - \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) \mathbb{B}_{12}(\epsilon)$. Using a formula similar to (A.10), and recalling the definition of $\mathbb{K}$ in (2.3.4) we have

$$\begin{aligned} \lim_{\epsilon \rightarrow \infty} \mathbb{B}_{11}^{-1}(\epsilon) &= \underline{\Sigma}_{11}^{-1} - \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \left(\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right)^{-1} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \\ &= \mathbb{K}. \end{aligned} \quad (\text{A.13})$$

Next, observe that

$$\begin{aligned} \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) &= \underline{\Sigma}_{21} \mathbb{B}_{11}^{-1}(\epsilon) + \epsilon \underline{X}^{(2)} \underline{X}^{(1)T} \\ &\quad \times \left(\underline{\Sigma}_{11} + \epsilon \underline{X}^{(1)} \underline{X}^{(1)T} \right)^{-1}. \end{aligned} \quad (\text{A.14})$$

Again, by Sallas and Harville (1981),

$$\begin{aligned} \lim_{\epsilon \rightarrow \infty} \epsilon \underline{X}^{(1)T} \left(\underline{\Sigma}_{11} + \epsilon \underline{X}^{(1)} \underline{X}^{(1)T} \right)^{-1} &= \left(\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right)^{-1} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1}. \end{aligned} \quad (\text{A.15})$$

From (A.13) - (A.15) and (2.3.5), it follows that

$$\begin{aligned} \lim_{\epsilon \rightarrow \infty} \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) &= \underline{\Sigma}_{21} \mathbb{K} + \underline{X}^{(2)} \left(\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)} \right)^{-1} \underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \\ &= \mathbb{M}. \end{aligned} \quad (\text{A.16})$$

Now, note that

$$\begin{aligned}
 & \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) \mathbb{B}_{12}(\epsilon) \\
 &= \Sigma_{21} \mathbb{B}_{11}^{-1}(\epsilon) \Sigma_{12} + \epsilon \Sigma_{21} \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \mathbf{X}^{(2)T} \\
 &\quad + \epsilon \mathbf{X}^{(2)} \mathbf{X}^{(1)T} \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \Sigma_{12} \\
 &\quad + \epsilon^2 \mathbf{X}^{(2)} \mathbf{X}^{(1)T} \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \mathbf{X}^{(2)T}.
 \end{aligned} \tag{A.17}$$

Now since $\mathbf{X}^{(1)}$ is of full column rank, there exists some matrix $\mathbb{F}$ s.t. $\mathbf{X}^{(2)} = \mathbb{F} \mathbf{X}^{(1)}$. Then

$$\begin{aligned}
 & \epsilon^2 \mathbf{X}^{(2)} \mathbf{X}^{(1)T} \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \mathbf{X}^{(2)T} \\
 &= \epsilon \mathbb{F} \left(\epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right) \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \mathbf{X}^{(2)T} \\
 &= \epsilon \mathbb{F} \mathbf{X}^{(1)} \mathbf{X}^{(2)T} - \mathbb{F} \Sigma_{11} \left\{ \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \right\} \mathbf{X}^{(2)T} \\
 &= \epsilon \mathbf{X}^{(2)} \mathbf{X}^{(2)T} - \mathbb{F} \Sigma_{11} \left\{ \left(\Sigma_{11} + \epsilon \mathbf{X}^{(1)} \mathbf{X}^{(1)T} \right)^{-1} \mathbf{X}^{(1)} \right\} \mathbf{X}^{(2)T}.
 \end{aligned} \tag{A.18}$$

From (A.17) and (A.18), it follows that

$$\begin{aligned}
 \mathbb{B}_{22.1}(\epsilon) &= \mathbb{B}_{22}(\epsilon) - \mathbb{B}_{21}(\epsilon) \mathbb{B}_{11}^{-1}(\epsilon) \mathbb{B}_{12}(\epsilon) \\
 &= \Sigma_{22} - \Sigma_{21} \mathbb{B}_{11}^{-1}(\epsilon) \Sigma_{12}
 \end{aligned}$$

$$\begin{aligned}
& - \Sigma_{21} \left\{ \left(\Sigma_{11} + \epsilon X^{(1)} X^{(1)T} \right)^{-1} X^{(1)} \epsilon \right\} X^{(2)T} \\
& - X^{(2)} \left\{ \epsilon X^{(1)T} \left(\Sigma_{11} + \epsilon X^{(1)} X^{(1)T} \right)^{-1} \right\} \Sigma_{12} \\
& + F \Sigma_{11} \left\{ \left(\Sigma_{11} + \epsilon X^{(1)} X^{(1)T} \right)^{-1} X^{(1)} \epsilon \right\} X^{(2)T}.
\end{aligned} \tag{A.19}$$

Now using (A.13) and (A.15) we have from (A.19) that

$$\begin{aligned}
& \lim_{\epsilon \rightarrow \infty} B_{22.1}(\epsilon) \\
& = \Sigma_{22} - \Sigma_{21} K \Sigma_{12} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(2)T} \\
& - X^{(2)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(1)T} \Sigma_{11}^{-1} \Sigma_{12} \\
& + F \Sigma_{11} \Sigma_{11}^{-1} X^{(1)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(2)T} \\
& = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} + \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\
& \times X^{(1)T} \Sigma_{11}^{-1} \Sigma_{12} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(2)T} \\
& - X^{(2)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(1)T} \Sigma_{11}^{-1} \Sigma_{12} \\
& + X^{(2)} \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} X^{(2)T}, \text{ since } X^{(2)} = F X^{(1)}. \\
& = \Sigma_{22.1} + \left(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \right) \left(X^{(1)T} \Sigma_{11}^{-1} X^{(1)} \right)^{-1} \\
& \times \left(X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \right)^T, \text{ after some rearrangements} \\
& = G \quad \text{using (2.3.6)}.
\end{aligned} \tag{A.20}$$

Note that G is p.d. Now from (A.10), (A.12), (A.13) and (A.20), one gets

$$\underline{y}^T \underline{Q} \underline{y} = \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} + \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right)^T \underline{G}^{-1} \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right). \quad (\text{A.21})$$

Now from (A.6), (A.9) and (A.21) it follows that

$$\begin{aligned} & f(\underline{y}^{(2)} | \underline{\lambda}, \underline{y}^{(1)}) \\ & \propto |\underline{G}|^{-\frac{1}{2}} \left(\underline{a}_0 + \sum_{i=1}^t \underline{a}_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right)^{-\frac{1}{2}(N_T - n_T)} \\ & \times \left\{ 1 + \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right)^T \underline{G}^{-1} \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right) \right. \\ & \left. \div \left(\underline{a}_0 + \sum_{i=1}^t \underline{a}_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right) \right\}^{-\frac{1}{2}(N_T + \sum_{i=0}^t g_i - p)} . \end{aligned} \quad (\text{A.22})$$

Writing $\left(n_T + \sum_{i=0}^t g_i - p \right)^{-1} \left(\underline{a}_0 + \sum_{i=1}^t \underline{a}_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right) \underline{G} = \underline{\Omega}$,

it follows from (A.22) that

$$\begin{aligned} & f(\underline{y}^{(2)} | \underline{\lambda}, \underline{y}^{(1)}) \\ & \propto |\underline{\Omega}|^{-\frac{1}{2}} \left\{ n_T + \sum_{i=0}^t g_i - p + \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right)^T \underline{\Omega}^{-1} \right. \\ & \left. \times \left(\underline{y}^{(2)} - \underline{M} \underline{y}^{(1)} \right) \right\}^{-\frac{1}{2}(n_T + \sum_{i=0}^t g_i - p + N_T - n_T)} . \end{aligned} \quad (\text{A.23})$$

From (A.23) and (2.3.2), it follows that the conditional distribution of $\underline{Y}^{(2)}$ given $\underline{Y}^{(1)} = \underline{y}^{(1)}$ and $\underline{\Lambda} = \underline{\lambda}$ is $(N_T - n_T)$ -variate t -distribution with d.f.

$n_T + \sum_{i=0}^t g_i - p$, location parameter $\underline{y}^{(1)}$ and scale

parameter $\underline{\Omega} = \left(n_T + \sum_{i=0}^t g_i - p \right)^{-1} \left(a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right) \underline{G}$.

Also, since $f(\lambda | \underline{y}^{(1)}) = \frac{h(\underline{y}^{(1)}, \lambda)}{f(\underline{y}^{(1)})} \propto h(\underline{y}^{(1)}, \lambda)$, it follows

from (A.5) that

$$f(\lambda | \underline{y}^{(1)}) \propto |\underline{\Sigma}_{11}|^{-\frac{1}{2}} |\underline{X}^{(1)T} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)}|^{-\frac{1}{2}} \left[\prod_{i=1}^t \lambda_i^{\frac{1}{2} g_i - 1} \right] \\ \times \left(a_0 + \sum_{i=1}^t a_i \lambda_i + \underline{y}^{(1)T} \underline{K} \underline{y}^{(1)} \right)^{-\frac{1}{2} (n_T + \sum_{i=0}^t g_i - p)}.$$

APPENDIX B
AN INDEPENDENCE RESULT IN A FAMILY OF
ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

Lemma B.1. Consider a family of distributions $\mathcal{P} =$

$\{\mathfrak{s}_f(\underline{X}\underline{\beta}, \sigma^2\underline{\Sigma}) : \underline{\beta} \in \mathbb{R}^p, \sigma > 0\}$ where $\mathfrak{s}_f(\cdot, \cdot)$ is as defined by (3.3.18), $\underline{X}$ ($n \times p$) is a known matrix of rank p and $\underline{\Sigma}$ ($n \times n$) is a known p.d. matrix. Then for a random variable $\underline{Y}$ ($n \times 1$) with distribution $\mathfrak{s}_f(\underline{X}\underline{\beta}, \sigma^2\underline{\Sigma})$, $\mathfrak{s}(\underline{Y}) = \begin{pmatrix} s_1(\underline{Y}) \\ \mathfrak{s}_2(\underline{Y}) \end{pmatrix}$ and $\mathfrak{z}(\underline{Y})$ are independently distributed where $s_1(\underline{Y}) = (\underline{Y}^T \underline{Q} \underline{Y})^{\frac{1}{2}}$, $\mathfrak{s}_2(\underline{Y}) = (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1} \underline{Y}$, $\mathfrak{z}(\underline{Y}) = \underline{Q} \underline{Y} / s_1(\underline{Y})$ and $\underline{Q} = \underline{\Sigma}^{-1} - \underline{\Sigma}^{-1} \underline{X} (\underline{X}^T \underline{\Sigma}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Sigma}^{-1}$.

Proof of Lemma B.1. To prove this lemma, we need to verify the conditions of Proposition 7.19 of Eaton (1983). First we will note that $\mathfrak{s}(\underline{Y})$ is sufficient for the family of distributions $\mathcal{P}$. Note that $\underline{Y}$ has pdf

$$f(\underline{y}; \underline{\beta}, \sigma) \propto f((\underline{y} - \underline{X}\underline{\beta})^T \underline{\Sigma}^{-1} (\underline{y} - \underline{X}\underline{\beta}) / \sigma^2) \quad (\text{B.1})$$

and

$$\begin{aligned} & (\underline{y} - \underline{X}\underline{\beta})^T \underline{\Sigma}^{-1} (\underline{y} - \underline{X}\underline{\beta}) / \sigma^2 \\ &= \underline{y}^T \underline{\Sigma}^{-1} \underline{y} / \sigma^2 - 2 \underline{\beta}^T \underline{X}^T \underline{\Sigma}^{-1} \underline{y} / \sigma^2 + \underline{\beta}^T \underline{X}^T \underline{\Sigma}^{-1} \underline{X} \underline{\beta} / \sigma^2. \end{aligned} \quad (\text{B.2})$$

Recall from (3.3.16) - (3.3.18) that f is assumed to be known. From (B.1) and (B.2), it follows that $\begin{pmatrix} Y^T \Sigma^{-1} Y \\ X^T \Sigma^{-1} Y \end{pmatrix}$ is sufficient for $\mathcal{P}$. But

$$Y^T Q Y = Y^T \Sigma^{-1} Y - (X^T \Sigma^{-1} Y)^T (X^T \Sigma^{-1} X)^{-1} (X^T \Sigma^{-1} Y)$$

gives that $\begin{pmatrix} Y^T \Sigma^{-1} Y \\ X^T \Sigma^{-1} Y \end{pmatrix}$ and $s(Y)$ are one-to-one functions of each other. So $s(Y)$ is a sufficient statistic for the family $\mathcal{P}$.

Let $\mathfrak{X} = R^n - \mathcal{M}(X)$ where $\mathcal{M}(X) = \{X\beta: \beta \in R^P\}$ is the vector space generated by the columns of the matrix X . Also, let $\mathcal{A} = \{B \in \mathfrak{B}^n: B \subseteq \mathfrak{X}\}$ where $\mathfrak{B}^n$ is the product σ -algebra in R^n . We have chosen $\mathfrak{X}$ in this way so that for $y \in \mathfrak{X}$, $y^T Q y > 0$ and $z(y)$ is defined. Note that $P_{\mathfrak{S}_f}[Y \in \mathfrak{X}] = 1$.

Also denote the Cartesian product of R^+ , the positive part of the real line, and R^P by $\mathcal{Y}$, i.e., $\mathcal{Y} = R^+ \times R^P$, and the σ -algebra of the subsets of $\mathcal{Y}$ by $\mathcal{C}_1 = \{C \in \mathfrak{B}^{P+1}: C \subseteq \mathcal{Y}\}$.

Now, consider the group $\mathfrak{G} = (G, \circ)$ where the set G is given by $G = \left\{ g \mid g = \begin{pmatrix} d \\ X\beta \end{pmatrix} \in R^+ \times \mathcal{M}(X) \right\}$ and $\circ$ is a binary operation which we define below. For $g_i = \begin{pmatrix} d_i \\ X\beta_i \end{pmatrix} \in G$, $i = 1, 2$, we define $g_1 \circ g_2 = \begin{pmatrix} d_1 d_2 \\ X(\beta_1 + d_1 \beta_2) \end{pmatrix}$. With this definition, $\mathfrak{G}$ is a group with $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ as the identity element and $(g)^{-1} = \begin{pmatrix} d^{-1} \\ -d^{-1} X\beta \end{pmatrix} \in G$ as the inverse element of $g \in G$.

For $\underline{y} \in \mathfrak{X}$, G acts on $\mathfrak{X}$ from the left by $\underline{g}(\underline{y}) = d\underline{y} + \underline{X}\underline{\beta} \in \mathfrak{X}$. Note that the result of G acting on $\mathfrak{X}$ from the left is equivalent to the group of transformations

$$\mathfrak{G}_2 = \{ \underline{g}_{\underline{\beta}, d}, \underline{\beta} \in \mathbb{R}^P, d > 0: \underline{g}_{\underline{\beta}, d}(\underline{y}) = d\underline{y} + \underline{X}\underline{\beta} \}$$

given by (3.5.15).

Consider another group $\tilde{\mathfrak{G}} = (\tilde{G}, *)$ where $\tilde{G} = \{ \underline{\gamma} | \underline{\gamma} = \begin{pmatrix} d \\ \underline{\beta} \end{pmatrix} \in \mathbb{R}^+ \times \mathbb{R}^P \}$ and $*$ is a binary operation which we define for $\underline{\gamma}_i = \begin{pmatrix} d_i \\ \underline{\beta}_i \end{pmatrix} \in \tilde{G}$, $i = 1, 2$ by $\underline{\gamma}_1 * \underline{\gamma}_2 = \begin{pmatrix} d_1 d_2 \\ d_1 \underline{\beta}_2 + \underline{\beta}_1 \end{pmatrix}$. It is easy to verify that $\tilde{\mathfrak{G}}$ is a group with $\begin{pmatrix} 1 \\ \underline{0} \end{pmatrix} \in \tilde{G}$ as the identity element and $(\underline{\gamma})^{-1} = \begin{pmatrix} d^{-1} \\ -d^{-1}\underline{\beta} \end{pmatrix} \in \tilde{G}$ as the inverse element of $\underline{\gamma} \in \tilde{G}$.

Now, we will show that $\tilde{G}$ is a homomorphic image of G . Define the function η from G onto $\tilde{G}$ by

$$\eta(\underline{g}) = \begin{pmatrix} d \\ \underline{\beta} \end{pmatrix} \in \tilde{G}$$

where $\underline{g} = \begin{pmatrix} d \\ \underline{X}\underline{\beta} \end{pmatrix} \in G$. It is easy to verify that for $\underline{g}$, $\underline{g}_1$ and $\underline{g}_2 \in G$, $\eta(\underline{g}_1 \circ \underline{g}_2) = \eta(\underline{g}_1) * \eta(\underline{g}_2)$ and $\eta(\underline{g}^{-1}) = (\eta(\underline{g}))^{-1}$. Then the function η is a homomorphism from G onto $\tilde{G}$ and $\tilde{G}$ is a homomorphic image of G .

For $\underline{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \in \mathcal{Y}$, $\tilde{G}$ acts on $\mathcal{Y}$ from the left by

$$\underline{\gamma}(\underline{s}) = \begin{pmatrix} d s_1 \\ d s_2 + \underline{\beta} \end{pmatrix} \in \mathcal{Y}.$$

Sometimes we will say $\tilde{G}$ a group without explicitly referring to the group operation.

Note that $\tilde{G}$ acts measurably on $(\mathcal{Y}, \mathcal{C}_1)$ and $\tilde{G}$ is a locally compact and σ -compact topological group (with the usual topology in $\mathbb{R}^{p+1}$) with σ -algebra $\mathcal{C}_1$. Also, the mapping $(\gamma, \underline{s}) \rightarrow \gamma(\underline{s}) = \begin{pmatrix} ds_1 \\ ds_2 + \underline{\beta} \end{pmatrix}$ from $(\tilde{G} \times \mathcal{Y}, \mathcal{C}_1 \times \mathcal{C}_1)$ to $(\mathcal{Y}, \mathcal{C}_1)$ is measurable. Here $\mathcal{C}_1 \times \mathcal{C}_1$ is a product σ -algebra of $\mathcal{C}_1$ with itself. So the conditions of Proposition 7.17 of Eaton (1983) are satisfied for $\tilde{G}$ and the measurable space $(\mathcal{Y}, \mathcal{C}_1)$. Also note that $\tilde{G}$ acts transitively on $\mathcal{Y}$. Since

$$\underline{s}(\underline{g}(Y)) = \underline{s}(dY + \underline{X}\underline{\beta}) = \begin{pmatrix} ds_1(Y) \\ ds_2(Y) + \underline{\beta} \end{pmatrix} = \gamma(\underline{s}(Y)),$$

so $\underline{s}$ on $\mathcal{X}$ to $\mathcal{Y}$ is measurable and equivariant.

It is easy to check that

$$\underline{Q}(dY + \underline{X}\underline{\beta}) = d\underline{Q}Y \quad (\text{B.3})$$

and

$$s_1(dY + \underline{X}\underline{\beta}) = ds_1(Y). \quad (\text{B.4})$$

The definition of $\underline{z}(Y)$ and (B.3) and (B.4) imply that $\underline{z}(dY + \underline{X}\underline{\beta}) = \underline{z}(Y)$. Hence $\underline{z}(Y)$ is a G -invariant function.

Moreover, if $Y \sim \mathcal{E}_f(\underline{Q}, \underline{\Sigma})$, then the distributions of the family of random variables $\{\underline{g}(Y) = dY + \underline{X}\underline{\beta} : \underline{g} \in G\}$ is given by the family of distributions $\mathcal{P}$. So all the conditions of Proposition 7.19 of Eaton (1983) are verified. The independence of $\underline{s}(Y)$ and $\underline{z}(Y)$ follows.

BIBLIOGRAPHY

- Basu, D. (1955). On statistics independent of a complete sufficient statistic. Sankhyā, 15, 377-380.
- Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error components model for prediction of county crop areas using survey and satellite data. J. Amer. Statist. Assoc., 83, 28-36.
- Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, 2nd Edition. Springer-Verlag, New York.
- Brackstone, G.J. (1987). Small area data: policy issues and technical challenges. Small Area Statistics: An International Symposium. Eds.: R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh. Wiley, New York, pp. 3-20.
- Carter, G., and Rolph, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. J. Amer. Statist. Assoc., 69, 880-885.
- Choudhry, G.H., and Rao, J.N.K. (1988). Evaluation of small area estimators: an empirical study. Unpublished manuscript.
- Cochran, W.G. (1939). The use of analysis of variance in enumeration by sampling. J. Amer. Statist. Assoc., 34, 492-510.
- Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of population. Ann. Math. Statist., 17, 164-177.
- Datta, G.S., and Ghosh, M. (1988). Minimum risk equivariant estimators of percentiles in location-scale families of distributions. Cal. Statist. Assoc. Bull., 37, 201-207.

- Dempster, A.P., Rubin, D.B., and Tsutakawa, R.K. (1981). Estimation in covariance components model. J. Amer. Statist. Assoc., 76, 341-353.
- Eaton, M.L. (1983). Multivariate Analysis: A Vector Space Approach. Wiley, New York.
- Ericksen, E.P. (1974). A regression method for estimating population of local areas. J. Amer. Statist. Assoc., 69, 867-875.
- Ericson, W.A. (1969). Subjective Bayesian models in sampling finite populations (with discussions). J. R. Statist. Soc., B, 31, 195-233.
- Faith, R.E. (1978). Minimax Bayes and point estimators of a multivariate normal mean. J. Mult. Anal., 8, 372-379.
- Fay, R.E., and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. J. Amer. Statist. Assoc., 74, 269-277.
- Ferguson, T.S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.
- Fleiss, J.L. (1986). Analysis of data from multiclinic trials. Controlled Clinical Trials, 7, 267-275.
- Fuller, W.A. and Harter, R.M. (1987). The multivariate components of variance model for small area estimation. Small Area Statistics: An International Symposium. Eds.: R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh. Wiley, New York, pp. 103-123.
- Ghosh, M. (1989). Hierarchical and empirical Bayes multivariate estimation. Tech. Report No. 330, Dept. of Statistics, University of Florida.
- Ghosh, M., and Lahiri, P. (1987a). Robust empirical Bayes estimation of means from stratified samples. J. Amer. Statist. Assoc., 82, 1153-1162.
- Ghosh, M. and Lahiri, P. (1987b). Robust empirical Bayes estimation of variances from stratified samples. Sankhyā, B, 49, 78-89.

- Ghosh, M., and Lahiri, P. (1988). Bayes and empirical Bayes analysis in multistage sampling. Statistical Decision Theory and Related Topics, V4, 1. Eds.: S.S. Gupta and J.O. Berger. Springer-Verlag, New York, pp. 195-212.
- Ghosh, M., and Lahiri, P. (in press). A hierarchical Bayes approach to small area estimation with auxiliary information. Proceedings of the Joint Indo-U.S. Workshop on Bayesian Inference in Statistics and Econometrics. Eds.: P.K. Goel and N. Iyenger. Springer-Verlag, New York.
- Ghosh, M., Lahiri, P., and Tiwari, R.C. (1989). Nonparametric empirical Bayes estimation of the distribution function and the mean. Commun. Statist.: Theory and Methods, 18, 121-146.
- Ghosh, M., and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. J. Amer. Statist. Assoc., 81, 1058-1062.
- Ghosh, M., and Rao, J.N.K. (1990). Small area estimation. Under preparation.
- Gianola, D., and Fernando, R.L. (1986). Bayesian methods in animal breeding theory. J. Anim. Sci., 63, 217-244.
- Hartigan, J.A. (1969). Linear Bayes methods. J. R. Statist. Soc. B, 31, 446-454.
- Hartley, H.O., and Rao, J.N.K. (1968). A new estimation theory for sample surveys. Biometrika, 55, 547-555.
- Harville, D.A. (1976). Extension of Gauss-Markov theorem to include the estimation of random effects. Ann. Statist., 4, 384-395.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc., 72, 320-340.
- Harville, D.A. (1985). Decomposition of prediction error. J. Amer. Statist. Assoc., 80, 132-138.
- Harville, D.A. (1988). Mixed-model methodology: theoretical justifications and future directions. Proceedings of the Statistical Computing Section of the ASA, 41-49.

- Harville, D.A. (in press). BLUP and beyond. Advances in Statistical Methods for Genetic Improvement of Livestock. Eds.: D. Gianola and K. Hammond. Springer-Verlag, New York.
- Harville, D.A., and Fenech, A.P. (1985). Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model. Biometrics, 41, 137-152.
- Harville, D.A., and Jeske, D.R. (1989). Mean squared error of estimation and prediction under a general linear model. Tech. Report No. 89-37, Statistical Laboratory, Iowa State University.
- Henderson, C.R. (1953). Estimation of variance and covariance components. Biometrics, 9, 226-252.
- Henderson, C.R. (1963). Selection index and expected genetic advance. Statistical Genetics and Plant Breeding. Eds.: W.D. Hanson and M.F. Robinson. Publication 982, 141-163. NAS-NRC, Washington, D.C.
- Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. Biometrics, 31, 423-447.
- Hill, B. (1965). Inference about variance components in one-way model. J. Amer. Statist. Assoc., 60, 806-825.
- Hill, B. (1977). Exact and approximate Bayesian solutions for inference about variance components and multivariate inadmissibility. New Developments in the Application of Bayesian Methods. Eds.: A. Aykac and C. Brumat. North-Holland, Amsterdam, pp. 129-152.
- Hill, B. (1980). Robust analysis of the random model and weighted least squares regression. Evaluation of Economic Models. Academic Press, New York, pp. 197-217.
- Hwang, J.T. (1985). Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions. Ann. Statist., 13, 295-314.
- Jeske, D.R., and Harville, D.A. (1987). Prediction interval procedures and (fixed-effects) confidence interval procedures for mixed linear models. Commun. Statist.: Theory and Methods, 17, 1053-1087.

- Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J. Amer. Statist. Assoc., 79, 853-862.
- Kelker, D. (1970). Distribution theory for spherical distribution and a location-scale parameter generalization. Sankhyā, A, 32, 419-430.
- Lahiri, P., and Rao, J.N.K. (1990). Robust estimation of mean square error of empirical best linear unbiased predictors. Preprint.
- Lahiri, P., and Tiwari, R.C. (in press). Nonparametric Bayes and empirical Bayes estimation of variances from stratified samples. Sankhya.
- Lehmann, E.L. (1983). Theory of Point Estimation. Wiley, New York.
- Lehmann, E.L. (1986). Testing Statistical Hypotheses, 2nd Edition. Wiley, New York.
- Lehmann, E.L., and Scheffe, H. (1950). Completeness, similar regions and unbiased estimation. Sankhyā, 10, 305-340.
- Lindley, D.V., and Smith, A.F.M. (1972). Bayes estimates for linear model (with discussion). J. R. Statist. Soc., B, 34, 1-41.
- Lui, K.J., and Cumberland, W.G. (1989). A Bayesian approach to small domain estimation. J. Off. Statist., 5, 143-156.
- Malec, D., and Sedransk, J. (1985). Bayesian inference for finite population parameters in multistage cluster sampling. J. Amer. Statist. Assoc., 80, 897-902.
- Morris, C. (1981). Parametric empirical Bayes confidence intervals. Scientific Inference, Data Analysis and Robustness. Eds.: G.E.P. Box, T. Leonard, and C.F.J. Wu. Academic Press, New York, pp. 25-50.
- Morris, C. (1983). Parametric empirical Bayes inference and applications. J. Amer. Statist. Assoc., 78, 47-65.
- Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared error of small area predictors. J. Amer. Statist. Assoc., 85, 163-171.

- Press, S.J. (1972). Applied Multivariate Analysis. Holt, Rinehart and Winston, New York.
- Rao, C.R. (1952). Some theorems on minimum variance estimation. Sankhyā, 12, 27-42.
- Rao, C.R. (1973). Linear Statistical Inference and Its Applications, 2nd Edition. Wiley, New York.
- Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. 3rd Berk. Symp. Math. Statist. Prob., V1, 157-164.
- Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. J. Amer. Statist. Assoc., 71, 657-664.
- Royall, R.M. (1979). Prediction models in small area estimation. Synthetic Estimates for Small Area. NIDA Research Monograph Series No. 24, pp. 63-87.
- Sallas, W.M., and Harville, D.A. (1981). Best linear recursive estimation for linear models. J. Amer. Statist. Assoc., 76, 860-869.
- Scott, A., and Smith, T.F.M. (1969). Estimation in multi-stage surveys. J. Amer. Statist. Assoc., 64, 830-840.
- Searle, S.R. (1971). Linear Models. Wiley, New York.
- Strawderman, W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist., 42, 385-388.
- Stroud, T.W.F. (1987). Bayes and empirical Bayes approaches to small area estimation. Small Area Statistics: An International Symposium. Eds.: R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh. Wiley, New York, pp. 124-137.
- Zellner, A. (1972). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

BIOGRAPHICAL SKETCH

Gauri Sankar Datta was born on January 4, 1962, in Midnapur, West Bengal, India. After graduating from Garbeta College, West Bengal, in 1980, he enrolled at RamaKrishna Mission Residential College, Narendrapur, a suburb of Calcutta. Upon receiving his Bachelor of Science degree with honors in statistics in 1983, he joined the Department of Statistics of Calcutta University where he received his Master of Science degree in statistics in 1986. He had gained some experience in teaching in R.K.M.R. College before coming to the University of Florida at Gainesville in the Spring of 1987. He expects to receive a degree of Doctor of Philosophy in statistics in August, 1990. During his time at the University of Florida, he was employed as a graduate consultant, performing consulting duties at the Institute of Food and Agriculture Sciences Statistics Unit. He was also employed as a teaching assistant in the Department of Statistics. In addition, he worked as a research assistant to Dr. Ramon C. Littell and Dr. Malay Ghosh.

He gained some valuable experience in computing while working with actual data in the United States Census Bureau where he accompanied Dr. Malay Ghosh as an ASA/NSF/Census Research Associate before his graduation. Upon graduation, he will join the Department of Statistics, University of Georgia, Athens.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Malay Ghosh

Malay Ghosh, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald H. Randles

Ronald H. Randles
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael A. DeLorenzo

Michael A. DeLorenzo
Associate Professor of Dairy
Science

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1990

Dean, Graduate School

UNIVERSITY OF FLORIDA



3 1262 08556 7252